Health risks associated with exposure to indoor air pollutants are heightened when biological vulnerability exists

indooraircartoon.com/2025/05/29/health-risks-associated-with-exposure-to-indoor-air-pollutants-are-heightened-when-biological-vulnerability-exists

29 May 2025

Indoor Air Cartoon Journal, May 2025, Volume 8, #166

[Cite as: Fadeyi MO (2025). Health risks associated with exposure to indoor air pollutants are heightened when biological vulnerability exists. *Indoor Air Cartoon Journal*, May 2025, Volume 8, #166.]

REFERENCES

Foley, P., and Riederer, P. (2000). Journal of Neurology, 247, II82-II94.
Jenner, P. (2003). Annals of Neurology, 53(S3), S26-S38.
Niu, X., et. al. (2021). Science of the Total Environment, 767, 144391.

Fictional Case Story (Audio – available online) – Part 1

Fictional Case Story (Audio – available online) – Part 2

Fictional Case Story (Audio – available online) – Part 3

Fictional Case Story (Audio – available online) – Part 4

……………………………….

There was a time in a country called Belinburg when indoor air quality (IAQ) was assessed using outdated methods that created the illusion of safety but failed to offer real protection. Professionals relied on time-weighted averages, spot sampling, and oversimplified assumptions that treated indoor environments as static, with predictable ventilation and uniform occupancy. Buildings were often certified as “safe” based on generalised metrics, with little regard for how people actually lived, moved, and breathed within those spaces.

This practice was not unique to Belinburg. Across the world, IAQ assessment frameworks failed to recognise exposure as a dynamic interaction, between emission sources, building characteristics, occupant behaviour, and biological vulnerability. In Belinburg, indoor environments frequently accumulated pollutants from everyday activities: cooking, incense burning, chemical sprays, and emissions from synthetic furnishings. Yet, as long as no single pollutant exceeded a short-term regulatory threshold, the environment was marked as compliant. Cumulative exposure, and the distinct vulnerabilities of older adults and individuals with chronic conditions, were ignored.

Even as artificial intelligence began transforming other sectors such as medicine and finance, its integration into IAQ assessment remained limited. Although real-time sensor data were increasingly available, the computational models remained static and disconnected from lived realities. Risk forecasts failed to reflect actual conditions.

In Belinburg, reform was slow to come, until a young woman, once shaped by her own misunderstanding, chose to act. She remembered her grandfather’s quiet wisdom: “A smaller flame, when focused with purpose, when shaped by care and informed by listening, could soften the hardest metal.” That insight, once overlooked, became the foundation for her work. She made a conscious decision to change the way IAQ was understood and practised. Her journey is the subject of this fiction story.

1……………………………….

In the sun-scorched corner of her grandfather’s backyard workshop in a city called Asamera in Belinburg, six-year-old Zahra stood on a splintered wooden crate, her bare feet coated with a thin layer of dust and heat. She grasped the iron handles of the bellows, each pump stiff and reluctant under her small hands. The rusted hinge creaked with each movement, but she pressed on, drawing air into the coals and watching the fire flare.

With every push, the flame hissed and spat like a creature waking from sleep. Sparks leapt into the shaft of sunlight streaming through the louvred window, their glow brief but brilliant. Her arms trembled, her breath came fast, and sweat rolled in rivulets down the side of her face. Still, she refused to stop. She believed that if she pushed hard enough, if she gave enough of herself, the fire would respond. She would be the one who made it roar.

Her grandfather, a man of slow gestures and measured words, stood to one side, shaping a horseshoe on the anvil with quiet, deliberate strikes. He had allowed Zahra into his sacred space not because he needed help, but because he had long ago learned that children must touch heat to understand fire. When she called out, “Grandpa, look how strong I am!” her voice rang with the certainty only children possess. She expected praise, perhaps even awe.

Grandpa paused, turned to look at her, and smiled with the tenderness that came from watching generations try to master the same lessons. He replied, “That is good, Zahra. Though sometimes, a smaller flame does the work better.” She barely heard him. His words brushed past her like a passing breeze, noticed but dismissed. She had no use for small flames. She wanted the flame that danced and growled, the flame that marked her effort, the flame that announced she was here.

Zahra did not understand then that the forge was not a place for proving strength. It was a place for shaping restraint. Her grandfather’s life had been built on reading temperature with his eyes, listening to metal with his hands, and knowing precisely when to stop. He had grown into a man who no longer admired loud flames but respected those that burned clean and steady. He said nothing more. His wisdom, like his fire, waited patiently.

The misbelief Zahra carried away from that day, equating value with effort, scale, and spectacle, followed her into the classrooms of her primary school, into the design competitions of her adolescence, and into the hollow corridors of acclaim. She came to believe that the worth of a solution was directly proportional to the scale of her effort, the volume of materials used, the amount of money acquired to execute it, and the time and energy invested. Whenever a problem was presented, Zahra instinctively began gathering tools, sketching grand ideas, and sourcing the most elaborate and impressive components she could find.

In Year Four, when her class was assigned to create a simple water filtration model, most of her classmates arrived with cut bottles, cotton pads, and labelled diagrams. Zahra entered with a wooden frame taller than her desk, fitted with multiple stages of gravel, charcoal, and solar-powered UV lights. Her teacher gasped. Her classmates marvelled.

The local paper sent a photographer. She was declared brilliant. No one asked about its cost. No one checked whether it could be built again. No one evaluated its effectiveness beyond appearance. Zahra never asked either. She believed that if the crowd clapped, the solution must have been right.

This belief hardened as she grew older. At sixteen, Zahra led her school’s entry into the national sustainable classroom design challenge. The brief called for low-energy cooling solutions to make learning environments more comfortable in regions with limited electricity. She proposed a solar-powered air-cooling unit combining copper coil evaporators, sensors programmed for temperature modulation, and motors salvaged from old treadmills.

She worked tirelessly, reducing her sleep to a few hours per night. She barked instructions, retested components repeatedly, and monitored output curves with obsessive precision. Her teachers were impressed by her determination, and her peers stepped back to allow her lead. The team won third place. The prototype was installed in a public school in East Asamera.

Three days after installation, the system failed. The pressure inside one of the copper coils exceeded tolerance, resulting in a crack that released the coolant. Teachers unplugged the device and placed a fan near the window instead. No one reported the failure. Zahra was not called to fix it.

A month later, she visited the school to conduct a follow-up. She expected a maintenance query, perhaps a request for additional funding. What she found was different. In the corner of the school's supply storage room, next to a broken globe and stacked filing cabinets, the cooling unit sat untouched. Her label—ZAHRA HAKIMI– PROJECT LEAD—was still written in bold black marker, now slightly smudged. The copper piping had oxidised to a dull brown. Dust gathered on the solar panel. The control screen was cracked.

She stood there for a long time, the sounds of children playing outside muffled by the concrete walls. Her eyes did not blink. She read her own name several times, the way one might stare at an old scar they had forgotten was there. For the first time in her young life, applause did not echo in her mind. There was only stillness.

The pride that had once surged within her felt strangely muted. It was not embarrassment she felt. Nor was it regret. It was something quieter, heavier, a slow erosion of certainty. She had built something beautiful, but no one had needed it. She had invested countless hours, pulled sleepless nights, used parts from three different cities, and created a machine that ultimately meant nothing to those it was meant to help.

The flaw her grandfather had seen was beginning to reveal itself. She realised that she had never once asked what problem the school was truly trying to solve. She had asked what she could build, not what they needed. She had focused on the fire, not the cooking. She had made it roar, but she had not fed anyone with it.

Her grandfather's words from a decade earlier returned to her now with surprising clarity. They no longer sounded like vague wisdom or quiet restraint. They sounded like the lesson she had ignored for too long. In the stale air of the storage room, surrounded by discarded things that once carried hope, Zahra finally understood. Perhaps the flame that burns the longest is the one that begins small and is never forgotten.

Two years had passed since the failure of her award-winning cooling prototype, and Zahra had neither forgotten the experience nor been able to explain what exactly had shifted in her since. She continued to earn high marks, to design with precision and ambition, and to lead school projects that dazzled judges and made headlines.

Yet beneath each accolade was a slow-growing sense of unease, one that she never spoke about aloud, but which crept into her drawings, her daydreams, and the way she paused a little longer before celebrating a finished build. Her classmates still admired her. Her teachers still praised her initiative. However, something within her had started to hollow, like the sound a cracked bell makes, resonant, but off-key.

When she received a full university scholarship to the country's top engineering school at the National University of Belinburg, the house filled with neighbours, relatives, and laughter. Her mother cooked Jollof rice with plantains and spiced meat. Her father stood straighter than she had ever seen him, wearing the agbada he reserved only for weddings and national holidays. "My daughter, the engineer," he said to the crowd, with a mix of pride and relief that only a man who had once set aside his own dreams could truly understand.

Neighbours offered prayers. A distant uncle promised to help with visa paperwork if she wanted to study abroad for a semester. Everyone believed she had reached the summit of success, the kind that not only changes a child's life but elevates an entire family's standing.

Yet Zahra hesitated. On the surface, the scholarship was everything she had worked for. The university had the most advanced fabrication labs in the country, a high student-to-faculty ratio, and partnerships with multinational firms that promised jobs before graduation. The brochure she received was printed on thick, glossy card with full-colour photographs of wind tunnels, robotic arms, and graduates in gowns. The headline on the back read: *We Build Tomorrow*.

That evening, long after the guests had left and the firewood smoke from celebratory cooking had faded into the air, Zahra slipped out quietly. The sky was turning amber, and the quiet hum of insects filled the silence. She walked alone to her grandfather's workshop, a place now layered with stillness and memory. The forge was cold. The bellows sat against the wall, its handles untouched in months. The anvil was covered in dust, and the hammer rested where her grandfather had left it before his last illness. She sat beside the bellows, her back against the stone wall, and closed her eyes.

In the silence, a memory returned. She saw a woman from one of her previous project visits, an older woman with thick, worn fingers and eyes that did not blink easily. Zahra had proudly demonstrated a high-efficiency stove prototype. The woman had smiled, reached out to touch the stainless-steel frame, and said, kindly and without judgement, "This looks expensive, my dear. What if it breaks? Will it still work next year?"

Zahra had not answered then. She had smiled politely, nodded vaguely, and moved on to the next house. She had not known what to say. Now, sitting beside the cold bellows, she realised she still did not have an answer. She had designed machines that solved problems her users never defined, in contexts she had only studied through lenses of abstraction. What she called "innovation" had often been a performance of intellect, not an act of service. The woman's question returned with sharpness and grace: Will it still work next year?

Zahra stood, walked slowly back to her room, and opened her laptop. She declined the scholarship and chose to accept the offer from the Belinburg Institute of Technology. During the university admissions period, she had searched through the directory of public universities and found a quiet programme with a plain title and few photographs—Sustainable Design and Value Systems at the Belinburg Institute of Technology. Its web page was minimal. No grand mission statement. No partnership logos. Just a simple sentence: We teach students to see clearly and design meaningfully.

No. I think I have been solving the wrong problem all along," she said. Her voice did not waver. It was quiet, yet every word landed with the weight of something she had never fully allowed herself to feel before—truth. She did not speak to defend her choice, nor to persuade him. She spoke because the illusion had finally lifted, and what stood behind it was clearer than anything the glossy brochures or prestigious rankings could offer.

Her whole life, she had poured herself into building things, ambitious things, complex things, without truly asking who they were for, or whether they would last. For years, applause had drowned out doubt. That was no longer the case. This moment was not about turning her back on a future. It was about walking towards one that actually made sense. The words left her lips not as resistance, but as release. It was not rebellion. It was clarity, for the first time.

Her father did not speak immediately. He remained seated, eyes fixed on the letter in his hand as though it might somehow rewrite itself. His shoulders, once lifted in pride, had slackened into something closer to disbelief. For a moment, Zahra thought he had not heard her. But then, very slowly, he looked up. His eyes, once bright with expectation, now held something more difficult, something quieter. He searched her face, perhaps looking for immaturity, for uncertainty, for any trace of doubt he could reason away. What he saw instead was calm. Not arrogance. Not impulse. Just a kind of grounded knowing that did not invite debate.

He exhaled, long and low, and set the letter down beside him. "You sound like your grandfather," he said, not with resentment, but with a weary sort of respect. "He always believed the fire was not what proved the metal. It was the shaping that came after." Then, after a long pause, he added, "Just promise me you will still build something." Zahra nodded, her throat tight and her eyes stinging. "I will," she said. "This time, something that lasts, and brings true value to those it is meant to serve.

2……………………………….

Zahar's experience in her undergraduate degree programme undid her. In her Year 2 of the programme, Zahra encountered a topic that would change her trajectory forever. In a module titled Health and Built Environments, she was introduced to the field of indoor air quality (IAQ). The instructor spoke not of gadgets, but of invisible burdens. Particles that triggered asthma. Volatile organic compounds that leaked from furniture. Carbon monoxide that slipped in unseen and silent.

Zahra learnt that what people breathed indoors was sometimes more dangerous than anything they encountered outside. She saw that IAQ could not be understood through gadgets alone. It demanded systems thinking, behavioural analysis, and ethical sensitivity. She began to see IAQ not as a subfield of engineering, but as a complex, interdisciplinary mirror of how people live.

She read about homes where incense was burned daily in poorly ventilated spaces. She learnt about kitchens where cleaning products were used without knowledge of the gases they released. She saw how low-cost composite materials in wardrobes and cupboards slowly released formaldehyde into the air children slept in. The particles were invisible. The damage was not.

Zahra found herself shaken. IAQ was not just about air. It was about inequality, invisibility, and slow violence. It required more than solutions—it required understanding. That was when she knew: she had not chosen the wrong programme. She had chosen to see.

Her assignment in the module was deceptively simple: “Propose a method to reduce indoor smoke exposure in peri-urban homes—without buying or building anything.” Zahra stared at the page for hours. It felt like a trick. How could one solve anything without constructing a solution? She was used to intervention through addition—more materials, more mechanisms, more brilliance. She had never been asked to solve through understanding.

She visited the homes of families living in the city’s inner districts, where corrugated roofs amplified heat and windows were mostly kept shut for privacy and dust control. She watched women position benches to catch cross-ventilation during cooking. She saw children fanning smoke out with folded cardboard sheets. Some families burnt citrus peels to clear out odours. Others angled mirrors to reflect light and improve visibility in smoke-filled kitchens. No one had been trained. No one had called these behaviours “design.” Yet they were adaptive, thoughtful, and rooted in context.

Zahra returned from each visit not with sketches of machines, but with stories. She documented routines, behaviours, and unintended innovations. She drew airflow paths over floor plans. She observed how long incense sticks burned and used the visible smoke to estimate how polluted the air became over time. She did not propose new devices or impose technical fixes. Instead, she focused on what people were already doing to reduce smoke, even in small, improvised ways such as propping open kitchen shutters with stones, angling mirrors to reflect light through hazy air, or placing damp cloths near cooking areas to trap soot.

Through conversations, she highlighted how simple, consistent habits, like placing incense further from sleeping areas or using low-emission fuels when available, could make a meaningful difference without requiring additional time, effort, or cost. Her report included airflow diagrams based on her observations, showing how smoke tended to accumulate in corners, bounce off hard surfaces, or linger when doors and windows were opened in certain sequences.

These diagrams were then adapted into illustrated guides that helped families understand how to adjust ventilation pathways to encourage faster smoke clearance. For example, the guides showed how opening opposite windows before cooking created a more effective cross-draught, and how partially closing a door could prevent smoke from drifting into bedrooms. By making these airflow patterns visible and understandable, the guides enabled families to make small layout or timing adjustments that reduced how long smoke stayed trapped indoors.

Zahra shared these illustrated guides in group sessions with the community, encouraging neighbours to exchange ideas and tailor the suggestions to their own routines. The guides also validated the local strategies people were already using, helping residents recognise the value in their own practices.

Zahra did not bring new inventions. She helped people see their own homes more clearly and supported them in turning small insights into lasting improvements that reduced exposure in ways that made sense for their lives. She wrote, "Perhaps innovation is not always invention. Perhaps it is recognition. Perhaps it begins when we learn to see what is already there and ask how to help it work better."

In her third year, Zahra began an internship with a large engineering consultancy. The firm was piloting an IAQ solution in high-rise public housing estates. The design involved imported sensors, cloud dashboards, and app-based alerts. It cost more than $400,000. The team presented charts, animations, and predictions. The ministry loved it and funded the project.

Zahra initially admired the elegance of it all. However, that changed when she was asked to review sensor data from a unit located above a busy street market. The dashboard showed smooth curves, minor fluctuations, and a green status for most indoor air parameters. Yet during her visit, she observed something very different.

The occupants cooked indoors with the windows shut due to outdoor smoke from meat stalls below. An incense stick burned continuously on the shelf. The family cleaned the floor with a strong aerosol disinfectant while she was there. Within ten minutes, her throat itched and her eyes watered. The sensor, however, still reported "normal conditions."

She raised the issue with the IAQ lead, pointing to the clear mismatch between her physical symptoms and the sensor readings displayed on the dashboard. Her throat had itched, her eyes had watered, and there had been a visible haze in the room. Yet the monitor remained calmly in the green zone, signalling "good" indoor air quality.

The lead responded with measured patience, explaining first that the dashboard was based on fifteen-minute rolling averages. Short-term spikes, such as those caused by aerosol sprays or incense, were often smoothed out or treated as statistical outliers that did not affect the average trend. Zahra nodded, but she knew that explanation barely scratched the surface.

He continued, acknowledging that the sensor unit installed in the flat was primarily configured to detect carbon dioxide, $PM_{2.5}$, and carbon monoxide. It had only limited capacity to capture volatile organic compounds (VOCs), especially the reactive ones released by incense or chemical cleaners. These compounds were often present in significant quantities during everyday activities, yet the device either registered them weakly or not at all. The problem, however, was not just the range of indoor air pollutants being measured.

The placement of the sensor mattered too. It had been installed near the ceiling in a hallway, selected for its proximity to a power source and its protection from tampering. It was a location that made logistical sense, but not a location that represented actual human exposure. The family cooked in a compact corner kitchen with no extractor fan.

The incense stick was positioned near the floor, close to where the children sat. The cleaning product was sprayed in quick bursts onto surfaces near breathing height. All of these events happened away from the sensor, and their effects did not travel upward or outward in time to be meaningfully captured by it.

With the windows shut due to the dense meat smoke outside, the flat had no active ventilation. Indoor air pollutants built up and lingered in microenvironments—kitchen corners, low sleeping areas, the play mat near the incense shelf. The indoor air was not uniformly mixed. It never had been. Yet the sensor, being a single-point device, could only offer data from where it was placed. It could not speak for the indoor air that Zahra had breathed, which had stung her lungs within minutes.

Then came the part that unsettled Zahra the most. The IAQ lead explained that the dashboard's summary status, those reassuring green and yellow icons, was not the raw output of the sensors. It was the result of a modelling layer embedded within the monitoring software.

This model processed the sensor data using a series of assumptions: that the flat had typical occupancy patterns, that ventilation was consistent across the space, and that pollutant concentrations were evenly distributed in the indoor environment. These assumptions were used to simplify the data for display, generate health risk labels, and reduce false alarms.

None of those assumptions matched the reality Zahra had just witnessed. The family's actual living patterns, their use of incense, their cleaning habits, and their ventilation behaviour all departed from the model's expectations. The model had filtered and translated the data to produce a clean narrative, one that made sense mathematically, but not biologically. According to the system, the indoor air in the flat was safe. According to Zahra's lungs, it was not.

She left the visit with a tight chest and a heavier mind. It was not just that the system had missed something. It was that it had reinterpreted the indoor environment to appear more manageable than it was. It had transformed a polluted experience into a smooth curve. The data had not lied, but it had been allowed to forget what Zahra's body had remembered.

Zahra felt a weight drop in her stomach. The more she reviewed the cases assigned to her, the more disturbed she became. In one flat, a child with asthma shared a confined two-room space with two adults who smoked indoors every evening. The smoke lingered for hours, trapped by closed windows and heavy curtains. The only ventilation came from a small oscillating fan, which simply stirred the pollutants without removing them.

In another case, a family used mosquito coils every night to ward off insects. The odour was sharp, but the real concern lay in the fine particles and volatile organic compounds released into the air as the coils burned slowly near sleeping children. Yet in both cases, the final indoor exposure scores generated by the IAQ dashboard fell within what the system labelled "acceptable" or "moderate".

The discrepancy was not just surprising; it was unsettling. According to the data outputs, there was no significant cause for concern. According to basic toxicology and public health principles, however, these children were being exposed daily to a combination of irritants and potential long-term hazards. Zahra dug deeper and discovered that the models used by the system were based on long-standing industry conventions. These included assumptions of constant emission rates, steady-state concentrations, and zone-averaged conditions across entire rooms or dwellings.

Occupancy was assumed to follow typical patterns, and ventilation was modelled as predictable and evenly distributed. The real-world complexity she had witnessed, short bursts of high emissions, closed windows due to outdoor pollution, pollutants released near the floor or sleeping spaces—was effectively smoothed out by the models.

The gap between lived experience and modelled exposure was not just a scientific oversight. It was a public health risk. It introduced a dangerous complacency into the system, suggesting safety where there was, in fact, vulnerability. Back in her dormitory, Zahra opened her notebook, sat at her desk, and wrote furiously.

Her thoughts were unfiltered and raw, colliding into one another with urgency. She began listing the dominant practices she had seen used in the field. Time-weighted averages. Multivariate mass balance models. One-off spot sampling. Zone-based monitoring with single-point sensors. All of them shared a common flaw: they were built on simplifications that ignored real life.

These frameworks treated buildings as homogenous boxes, with evenly distributed pollutants and stable emission behaviour. They treated human behaviour as predictable and ventilation as continuous and efficient. None of these assumptions held true in the flats she had visited. People moved from room to room. They opened windows when cooking but kept them closed at night. Children sat close to the floor where heavier pollutants settled. Cleaning sprays were used in concentrated bursts. Pollutants did not spread evenly, nor did they behave in linear ways.

She paused and wrote a question that would later define her doctoral work: “If life is not lived in averages, why do we model exposure that way?” The more she read, the more frustrated she became. Artificial intelligence and machine learning were already transforming fields as diverse as medicine, agriculture, transport, and finance. These systems could identify non-linear relationships, personalise outcomes, adapt in real time, and even predict behaviours based on contextual signals.

Yet IAQ assessment remained strikingly static. Despite the increasing availability of real-time sensor data, few tools existed to model pollutant dynamics in ways that reflected real-world complexity. There was almost no effort to forecast exposure based on how people actually used their spaces or to stratify risk based on biological susceptibility.

Zahra questioned why a food delivery app could learn her routine and anticipate her hunger, yet no IAQ system could alert an asthmatic child when their air quality was about to worsen due to an incense stick. She wondered why AI could recommend a movie tailored to her past preferences, but not help identify the dominant sources of harmful exposure in a specific flat, based on real-time patterns. The problem was not the lack of data. The problem was the lack of imagination, integration, and intent.

3………………………………

What troubled Zahra most, however, came during a review of case data involving an elderly woman with early-stage Parkinson’s disease who lived alone in a poorly ventilated flat. The home had been categorised as “low-risk” based on standard indoor air quality protocols. The home had passed its indoor air audit because no individual indoor air pollutant exceeded short-term exposure limits.

Yet Zahra’s field notes told a different story. The woman burned incense each evening in a sealed bedroom, used aerosol disinfectants daily, and seldom opened windows due to construction dust and heavy traffic pollution outside. There was no mechanical ventilation and no form of air purification.

The flat functioned as an enclosed microenvironment with multiple sources of pollutant emissions, repeated daily, over extended periods. The system focused on acute thresholds. Zahra saw something different. The woman’s exposure was cumulative, chronic, and occurring within a body already biologically compromised.

What made the case stand out to Zahra was not the mere presence of indoor air pollutants, but the biological context in which they were occurring. She knew that many residents were exposed to similar indoor sources, but this woman was in her late seventies, already living with a neurodegenerative condition, and spending nearly all her time indoors.

She had no one to intervene, no cross-ventilation to dilute exposures, and no awareness of the health effects of incense or spray use. Zahra began to consider whether it was not the indoor air pollutants alone, but the interaction between exposure and underlying biological vulnerability that the IAQ system had failed to register.

As she delved into recent literature, a disturbing pattern began to emerge. A growing number of studies had begun to link long-term exposure to pollutants such as fine particulate matter ($PM_{2.5}$), formaldehyde, nitrogen dioxide, and carbon monoxide with biological pathways involved in neurodegenerative diseases, including Parkinson’s.

These indoor air pollutants had been shown in experimental and epidemiological research to induce oxidative stress, mitochondrial dysfunction, chronic neuroinflammation, and protein misfolding—mechanisms that are implicated in the progression of Parkinson’s pathology.

Zahra also learnt from the literature that Parkinson’s disease was increasingly being recognised not only as a condition with genetic and age-related causes, but one in which environmental factors, especially indoor air pollution, played a role in accelerating onset or worsening progression.

While she understood that causality was complex and multifactorial, she also recognised that the current IAQ framework made no provision for individual susceptibility. It treated all occupants as physiologically equal. The system failed to ask whether an indoor air pollutant that might be tolerable for a healthy adult could still pose serious risks for someone or an older person with pre-existing neuroinflammation and reduced detoxification capacity.

The flat was not unsafe for everyone. However, it may have been unsafe for her. And that, Zahra realised, was precisely the problem. She began to question how many similar environments had been classified as “safe” under generalised standards, while slowly and silently intensifying health risks for biologically vulnerable individuals.

The IAQ assessment had not lied, but it had overlooked what mattered most. It assessed indoor air pollutants as isolated metrics, but not how they behaved in combination, over time, inside fragile human systems. For Zahra, this was not simply a gap in modelling practice. It was an ethical blind spot.

Health outcomes treated in hospitals, particularly progressive diseases like Parkinson’s, could, in many cases, be shaped by indoor environments that were never considered part of the problem. She saw the failure not only in the sensor or the threshold, but in the very philosophy of exposure assessment.

She also began to wonder whether artificial intelligence could be used not just to predict exposure patterns, but to model the interaction between indoor air pollutant profiles and biological vulnerability more accurately. Could machine learning, trained on real-time sensor data, building characteristics, and health profiles, help forecast risk for individuals with neurodegenerative predispositions? Could it be used to flag combinations of exposure and biological vulnerability that, while low on their own, became hazardous when experienced together?

Zahra saw potential for AI not only to enhance IAQ prediction, but to radically shift how health risk was defined, moving away from one-size-fits-all thresholds toward dynamic, biologically stratified forecasts that could inform both preventive care and targeted intervention.

That moment reframed everything. Parkinson’s was no longer simply a neurological concern best left to medicine. It was, at least in part, an environmental story written in routine indoor behaviours, accumulated through unnoticed emissions, and rendered invisible by models that ignored human diversity. What Zahra saw was not a single case of disease. She saw a system of assumptions that concealed risk instead of revealing it. And for the first time, she felt certain that understanding and correcting that system could help change the course of lives long before they reached a diagnosis.

In that moment, Zahra no longer viewed her research ideas as just an academic proposal. They became a necessity, something that needed to exist because lives and health outcomes depended on it. She decided she would pursue a PhD that aimed to dismantle the overly simplified frameworks that currently shaped IAQ assessment practice. Her goal was to rebuild the foundations of IAQ assessment around what people actually breathe, when they breathe it, and how their biological systems respond over time.

She began drafting her problem statement, shaping it as both a critique and a vision for what IAQ research and practice needed to become. She outlined how existing approaches failed to capture exposure variability in real time. Steady-state models neglected the erratic nature of human behaviour and building usage. Population-level thresholds also ignored the reality of vulnerable individuals, such as older adults, children with asthma, or people with chronic inflammatory conditions, whose exposure risk could not simply be averaged away.

She argued that mitigation strategies in use today remained overly general, recommending broad material substitutions or ventilation improvements that were not linked to actual emission events or specific health risks. She called for an AI-enhanced, biologically stratified, source-specific approach to IAQ assessment and intervention.

Her proposal was not simply to improve measurement accuracy. It was to close the gap between environmental data and lived human experience, so that IAQ systems could actually serve the people who breathed the air—not just satisfy regulatory requirements.

She submitted her PhD application with the problem statement at its core. Accompanying it were a clearly defined research aim, a well-structured set of research questions and objectives, and hypotheses grounded in what she had observed, experienced, and critically understood.

Her undergraduate record was strong. She had graduated with First Class honours, but it was the clarity and conviction of her proposal that distinguished her. Reviewers noted her rare combination of technical and scientific competence for a graduating student, paired with sharp observational insight and a deep sense of ethical responsibility. Her proposal was academically rigorous. It was driven by purpose.

For Zahra, the real achievement was not simply gaining acceptance into the programme. It was the deep sense that, for the first time, she had named the real problem, clearly, directly, and with intent. Below is an extract from Zahra's PhD proposal that captured the attention of the reviewers and anchored their decision.

"Across diverse urban environments, there exists a widening gap between current indoor air quality (IAQ) assessment practices and the expected assessment practices required to protect human health. Existing practices, often based on static averages, limited temporal resolution, and generalised assumptions, fail to accurately capture how people are exposed to pollutants under real-world conditions as expected.

This includes the dynamic interplay of building characteristics, emission events, occupant behaviours, and exposure durations. The expected performance level is one where indoor air pollutant exposure can be precisely characterised over time, across microenvironments, and stratified by biological risk, leading to timely, data-driven, and personalised interventions.

While this gap is observed across a range of settings, it becomes especially critical in densely polluted environments where high ambient pollution, limited indoor space, and intensive human activity can intensify indoor pollutant accumulation and elevate exposure risk. In less polluted environments, the gap remains relevant because chronic low-level exposures and undetected indoor sources can still present significant long-term health risks, particularly for individuals with underlying vulnerabilities.

Addressing this problem requires the development of more precise, dynamic, and risk-responsive IAQ assessment approaches. These approaches must move beyond generic estimations and instead characterise pollutant exposure as it actually occurs in lived environments. Such advancements are essential to enable health-protective strategies and promote equitable, evidence-based IAQ management.

The root of this problem lies in outdated IAQ observation and modelling methodologies. Industry practice continues to rely predominantly on time-weighted average (TWA) exposure estimations, one-off spot sampling, or zone-averaged sensor readings. These approaches oversimplify pollutant dynamics by assuming constant emission rates, uniform occupancy, and that indoor air pollutant concentrations remain in a steady-state equilibrium.

However, actual indoor environments are characterised by intermittent high-emission events (e.g., cooking, incense burning, cleaning with chemical sprays), variable ventilation patterns driven by weather and occupant behaviour, and the presence of low-level continuous emissions from synthetic materials. As a result, indoor air pollutant concentrations can fluctuate by orders of magnitude within hours, but current models fail to capture these fluctuations, leading to substantial underestimation of real cumulative exposure.

Furthermore, industry certification systems and regulatory frameworks do not adequately account for cumulative, long-term exposure to indoor neurotoxic pollutants such as $PM_{2.5}$, formaldehyde, benzene, nitrogen dioxide (NO_2), and carbon monoxide (CO), nor do they systematically integrate carbon dioxide (CO_2) as a proxy for inadequate ventilation that can amplify exposure risks.

This limitation is especially concerning given the growing body of epidemiological and toxicological evidence suggesting that prolonged exposure to these pollutants may contribute to biological mechanisms, such as chronic inflammation, oxidative stress, and mitochondrial dysfunction, that are implicated in the development and progression of neurodegenerative diseases like Parkinson’s. Despite this, IAQ assessments remain focused on short-term compliance metrics, neglecting the long-term health impacts of continuous low-dose exposure.

In the community, general awareness of these risks is limited. Occupants are often unaware of the role their daily activities and material choices play in degrading indoor air quality. Cultural practices, such as incense burning, fragrance use, or the installation of low-cost composite materials, persist without adequate information about their health implications. This is exacerbated by the absence of personalised, data-informed guidance from IAQ professionals. Occupants do not receive exposure feedback tailored to their space usage, vulnerability profile, or building characteristics.

This issue is hypothesised to be particularly critical for biologically vulnerable individuals, such as older adults, those with chronic inflammatory or metabolic conditions, or those with genetic predispositions to neurodegenerative diseases, who are believed to face disproportionately higher health risks from indoor pollutant exposure. However, current IAQ risk assessments are not stratified by biological susceptibility. The medical and building health domains remain largely siloed, limiting effective communication and coordinated intervention strategies.

Another key performance gap lies in mitigation strategy design. The current industry IAQ assessment practice still favours generalised ventilation recommendations or broad material substitutions rather than risk-based, source-specific strategies. This approach ignores the nuanced differences between indoor air pollutant types, emission persistence, and exposure pathways. For example, it treats gas stove emissions, pressed wood off-gassing, and fragrance aerosols as comparable IAQ risks, despite their distinct temporal profiles, chemical behaviours, and exposure consequences. Without reliable source attribution and exposure pathway modelling, interventions cannot be targeted, efficient, or scalable.

Meanwhile, the potential of artificial intelligence (AI) and machine learning to drive real-time IAQ diagnostics, predictive risk assessment, and scenario planning remains largely untapped in both industry and community contexts. Despite their demonstrated ability to model non-linear relationships among indoor air pollutant sources, building characteristics, ventilation dynamics, and occupant behaviours, machine learning-based exposure forecasting tools are rarely implemented in practice.

Although sensor data are increasingly available, the advanced computational frameworks required to transform this data into proactive, precision-based IAQ insights have yet to be meaningfully adopted. As a result, the built environment sector continues to lack the capacity to deliver dynamic, context-aware IAQ solutions aligned with evolving scientific knowledge.

This problem also creates an educational and training bottleneck. Current engineering, public health, and environmental curricula do not sufficiently prepare practitioners to work across the interface of exposure modelling, biological vulnerability assessment, and digital analytics. This impedes the development of a new generation of professionals capable of delivering context-responsive, health-driven IAQ strategies.

Bridging the gap in IAQ assessment practices demands integrated research that not only improves technical modelling and exposure analytics but also delivers actionable, context-specific, and biologically stratified interventions that are scalable across diverse building types and occupant profiles.

The research problem led to the following research questions: (i) How can long-term exposure to specific indoor air pollutants be accurately characterised and modelled in diverse indoor environments, considering dynamic factors such as building characteristics, occupant behaviours, ventilation patterns, and pollutant sources? (ii) How do long-term indoor air pollutant exposure profiles, as characterised by refined modelling approaches, interact with pre-existing health vulnerabilities, such as chronic inflammation, metabolic disorders, or genetic predispositions, to influence the development and progression of Parkinson's disease through mechanisms like oxidative stress, neuroinflammation, and mitochondrial dysfunction? (iii) What are the dominant indoor sources and exposure pathways contributing to long-term pollutant accumulation in residential and occupational settings, and how can this knowledge inform targeted, risk-based preventive strategies to protect populations most vulnerable to Parkinson's disease?

Each of the research questions determines the phase of the research study. The research questions and problems informed the following objectives of her PhD research: (i) To accurately characterise and model long-term exposure to specific indoor air pollutants in diverse indoor environments by accounting for dynamic factors such as building characteristics, occupant behaviours, ventilation patterns, and pollutant sources. (ii) To investigate how long-term indoor air pollutant exposure profiles, as characterised by refined modelling approaches, interact with pre-existing health vulnerabilities such as chronic inflammation, metabolic disorders, or genetic predispositions, in influencing the development and progression of Parkinson's disease through mechanisms including oxidative stress, neuroinflammation, and mitochondrial dysfunction. (iii) To identify the dominant indoor sources and exposure pathways contributing to long-term pollutant accumulation in residential and occupational settings, and to determine how this knowledge can inform the development of targeted, risk-based preventive strategies for protecting populations vulnerable to Parkinson's disease.

The research questions lead to the following research hypothesis. Hypothesis 1.1: A multivariate exposure model that incorporates dynamic indoor environmental factors, such as ventilation rate variability, building envelope characteristics, pollutant source profiles, and occupant behavioural patterns, will predict long-term indoor air pollutant exposure levels more accurately than static or average-based models. Hypothesis 1.2: Sensor-based AI modelling, when calibrated with building-specific and behavioural input data, will outperform traditional mass balance and time-weighted average models in estimating real-world chronic exposure profiles in diverse indoor environments.

Hypothesis 2.1: Individuals with pre-existing health vulnerabilities (e.g., chronic inflammation, mitochondrial dysfunction, or metabolic disorders) will show significantly greater biological responses, such as elevated oxidative stress markers and neuroinflammatory cytokines, to the same level of long-term indoor air pollutant exposure compared to healthy controls. Hypothesis 2.2: The interaction between modelled long-term indoor air pollutant exposure and pre-existing biological vulnerabilities will predict higher risk for Parkinson's disease development and progression than either factor alone.

Hypothesis 3.1: Combustion-related sources (e.g., cooking, heating, smoking) and off-gassing from indoor materials (e.g., furnishings, building products) are the dominant contributors to long-term accumulation of neurotoxic pollutants, such as $PM_{2.5}$, VOCs, and formaldehyde, in poorly ventilated residential and occupational indoor environments. Hypothesis 3.2: Risk-based preventive strategies tailored to high-emission sources and specific occupant vulnerabilities (e.g., targeted ventilation interventions, source removal, or behavioural guidelines) will achieve a significantly greater reduction in cumulative exposure risk compared to generic, non-targeted mitigation measures."

Zahra's PhD research was supervised by the famous and word renowned professor of engineering education, Professor Gulzar Adam. Below is an excerpt from Zahra's PhD thesis.

4………………………………

Research Methods

Methods for Research Question 1

Study Design and Site Selection

The study adopted a multi-phase, mixed-methods design that integrated extensive field-based data collection, continuous indoor air pollutant concentration monitoring, contextual variable measurement, and computational model development. A stratified sample of thirty indoor environments was selected to represent a cross-section of residential and occupational buildings. These included naturally ventilated and mechanically ventilated apartments, detached homes, offices, schools, and small commercial premises located across urban and peri-urban regions.

Stratification was based on criteria such as building age, floor plan configuration, occupant density, envelope airtightness, and the use of materials known to emit indoor air pollutants—particularly volatile organic compounds (VOCs), including formaldehyde.

The sampling framework deliberately incorporated buildings with differing floor levels, orientations, construction typologies, and operational characteristics, to reflect the full spectrum of environmental variability affecting long-term indoor air pollutant exposure. During an initial audit of each building, baseline data were recorded on geometry, room volume, surface finishes, occupant activity profiles, combustion-related behaviours, and air-conditioning systems.

Environmental Monitoring

Indoor air pollutant concentrations were monitored continuously over a period ranging from six to twelve months. This duration was chosen to ensure adequate capture of temporal variability, including daily, weekly, and seasonal fluctuations. The target pollutants, $PM_{2.5}$, VOCs (especially formaldehyde and benzene), carbon dioxide (CO_2), and carbon monoxide (CO), and nitrogen dioxide (NO_2), were selected based on their known or suspected neurotoxicity, persistence in indoor air, and potential role in the development or progression of neurodegenerative diseases such as Parkinson’s disease.

Two types of monitoring instruments were deployed at each site. Firstly, a network of calibrated low-cost sensors was installed, including optical particle counters (for $PM_{2.5}$), metal oxide sensors (for VOCs and NO_2), and non-dispersive infrared (NDIR) sensors (for CO_2). These sensors recorded data at five-minute intervals, enabling high temporal resolution.

Secondly, reference-grade instruments, such as gravimetric PM samplers and gas chromatography-based VOC detectors, were rotated periodically among the sites to provide calibration benchmarks and facilitate cross-validation. Environmental conditions, including temperature and relative humidity, were recorded in parallel, given their influence on chemical reaction rates and particle dynamics.

All sensors underwent pre-deployment calibration using laboratory standards and were recalibrated on-site using collocated reference measurements. Time synchronisation across instruments ensured accurate integration of pollutant concentration data with behavioural and contextual datasets.

Measurement of Ventilation and Building Parameters

Quantifying air exchange rates was essential to characterising pollutant ingress, retention, and removal. In mechanically ventilated buildings, airflow rates were measured using balometers and tracer gas decay techniques. For naturally ventilated environments, the CO_2 decay method was employed. This involved temporarily elevating indoor CO_2 concentrations, either via occupant presence or controlled release—and tracking the decay under steady-state conditions. Multiple measurements were taken throughout the year to capture variability induced by seasonal temperature gradients, wind speed, and window-opening behaviours.

Airtightness was assessed using blower door testing, in accordance with ASTM E779 standards. The resulting air leakage rates (reported as air changes per hour at 50 Pascals) were then used to estimate infiltration rates under natural pressure conditions. Additionally, digital building information models (BIM) were developed for each site, allowing accurate calculation of indoor volumes and the integration of structural parameters into the exposure models.

Occupant Behavioural Logging

To understand the influence of occupant behaviours on indoor air pollutant concentrations, both active and passive behavioural monitoring strategies were employed. Participants completed time-activity diaries over at least two weeks, documenting activities known to affect pollutant levels—such as cooking (including fuel type), cleaning (particularly with chemical sprays), smoking, burning incense, and use of air fresheners.

These self-reported data were supplemented with passive monitoring tools. Passive infrared (PIR) sensors captured occupancy patterns, while magnetic contact sensors monitored window and door status in real time. These behavioural datasets were used to inform the emission source terms in the mass balance model and served as critical features for training AI models. To reduce participant burden while capturing long-term patterns, behavioural data were collected in two intensive sampling campaigns spaced six months apart, thereby facilitating seasonal comparisons.

Development of Exposure Models

Three distinct exposure models were developed to estimate long-term exposure to indoor air pollutants. The first model, the time-weighted average (TWA) model, was constructed using the conventional formula:

$$E = \Sigma(C_i \times t_i) / \Sigma t_i$$

where E is the (average) estimated exposure over a defined period, C_i is the indoor air pollutant concentration at time interval i, and t_i is the time spent in the indoor environment during that interval. This model served as the baseline comparator.

The second model was a multivariate mass balance model. It incorporated source emission rates (estimated from literature or direct measurement), air exchange rates (measured as described), pollutant decay constants, and the volume of the indoor space. The model was governed by the differential equation:

$$dC/dt = P/V + (Q/V) \cdot (C_o - C) - kC$$

where C is the indoor air pollutant concentration, P is the generation rate, V is the indoor volume, Q is the ventilation rate, C_o is the outdoor concentration, and k is the decay constant. The model was solved numerically and validated using the empirical data. This equation can be rewritten as:

$$dC/dt = P/V + (Q/V) \cdot C_o - (Q/V) \cdot C - kC$$

when factorised and regrouped, the equation becomes

$$dC/dt = [P/V + (Q/V) \cdot C_o] - (Q/V + k) \cdot C$$

For context, this equation is mathematically equivalent to the simplified general differential form:

$$dC/dt = S - aC(t)$$

Where:

$$S = P/V + (Q/V) \cdot C_o \quad \text{and} \quad a = Q/V + k$$

In this context, S represents the combined source terms, including both indoor emissions and pollutant entry from outdoor air, while a represents the total first-order removal rate. The removal rate a comprises the air change rate (i.e., the ventilation rate per unit volume of the indoor environment, Q/V) and the first-order decay constant k, which accounts for additional removal processes such as filtration, surface deposition, chemical transformation, etc.

This differential formulation describes the instantaneous rate of change of the indoor air pollutant concentration (change in concentration at the moment) based solely on the current concentration C, assuming that the source and removal terms are known. While the equation itself does not require knowledge of the initial concentration C_0 to compute the rate at any given moment, numerical simulation of concentration dynamics over time requires specification of C_0 as a starting condition. In this study, the equation was solved numerically to simulate indoor air pollutant dynamics, and the model outputs were validated against empirical data collected from monitored indoor environments.

In cases where the system is assumed to have constant parameters—namely, constant source strength S, constant removal rate a, and a known initial concentration C_0, an analytical solution to the governing differential equation can be derived. The time-dependent solution for the indoor pollutant concentration is given by:

$$C(t) = S/a + (C_0 - S/a) \cdot e^{-at}$$

Differentiating this solution with respect to time yields:

$dC/dt = e^{-at} \cdot (S – aC_0)$

This expression describes how the rate of change of the indoor air pollutant concentration evolves explicitly over time (change in concentration over time interval), incorporating both the influence of the total source-to-removal ratio S/a and the initial concentration C_0. It highlights the transient behaviour of the system before reaching steady-state and is useful for validating numerical solutions under controlled conditions.

This expression for the rate of change, $dC/dt = e^{-at} \cdot (S – a \cdot C_0)$, provides insight into how the concentration evolves dynamically over time, particularly in the initial transient phase. It shows that the rate at which concentration increases or decreases is governed by the exponential decay of the difference between the source-driven equilibrium concentration S/a and the initial concentration C_0. When $C_0 < S/a$, the indoor concentration rises toward equilibrium, whereas when $C_0 > S/a$, the concentration declines. As time progresses, the exponential term diminishes, and the rate of change approaches zero, indicating that the concentration asymptotically stabilises at C = S/a.

However, while this dynamic solution provides valuable insight into how indoor air pollutant concentrations adjust toward equilibrium, it does not in itself quantify exposure over a defined period—something crucial for risk assessment. To estimate the average exposure experienced by occupants over a period of time T, the time-dependent concentration C(t) must be integrated over that interval. The appropriate formulation for this is the time-averaged exposure equation:

$E = (1/T) \times \int C(t)\,dt$, over the interval from 0 to T $\approx (1/T) \cdot \sum C_i \cdot t_i$ – *Using rectangular rule*

[***Note****: ≈* $(1 / T) \times \Sigma [C(t_{i-1}) + C(t_i)] \times t_i / 2$ *[Using Trapezoidal Rule] – when concentration changes significantly (>20-30%) between time points.*]

This equation computes the average indoor concentration over the period T, capturing the cumulative impact of fluctuating pollutant levels rather than just their value at a specific moment. It accounts for how long occupants are exposed and to what concentrations, making it directly relevant for health-based evaluations and regulatory comparisons.

To illustrate the use of this exposure model, a hypothetical scenario was considered in which indoor formaldehyde levels were monitored over an 8-hour workday. Concentration values were obtained hourly and assumed to remain constant within each 1-hour interval. Applying the rectangle method to approximate the integral, the total exposure load (area under the concentration–time curve) was computed by summing the product of concentration and time for each interval. This yielded a cumulative exposure of 151 $\mu g \cdot h/m^3$. Dividing by the total period (8 hours), the average exposure was found to be 18.875 $\mu g/m^3$.

In essence, the multivariate mass balance model can ultimately produce a result that resembles the Time-Weighted Average (TWA) model, but with a deeper explanation of what happens dynamically. This approach demonstrates how the analytical or numerical solution to the differential mass balance model can be operationalised to support exposure estimation.

Importantly, it shows that even in the absence of constant concentrations or steady-state conditions, reliable exposure estimates can still be obtained by integrating the output of the dynamic model. This capacity to translate transient concentration profiles into meaningful exposure metrics is essential for both acute and chronic health risk assessments, and further reinforces the value of dynamic modelling in real-world indoor air quality studies.

Although the numerical simulations employed in this study were based on solving the general form of the equation directly, the analytical solution and its derivative were used to confirm model consistency and provide theoretical insight into pollutant decay dynamics under idealised conditions.

The third modelling approach employed in this study was based on AI, specifically designed to capture complex, dynamic relationships between indoor environmental conditions and long-term pollutant exposure. This AI-based model was implemented as a computer programme, a set of coded instructions, that was run on a secure university server. This programme processed indoor environmental data and used machine learning algorithms to predict how much pollution people were exposed to over time inside buildings.

Multiple supervised machine learning techniques were tested, including random forest regression, support vector machines (SVMs), and long short-term memory (LSTM) neural networks. The LSTM models were particularly well-suited for handling time-series data and learning patterns that evolve over time.

The algorithms were trained using a multivariate time-series dataset collected from thirty buildings over 6 to 12 months. Inputs to the model included high-resolution sensor readings of indoor pollutant concentrations ($PM_{2.5}$, VOCs, CO, and CO_2), real-time ventilation data (derived from sensor-based window status and CO_2 decay), occupancy patterns (recorded via PIR motion sensors), temperature, humidity, and activity logs reporting pollutant-generating behaviours (e.g., cooking, cleaning, and chemical use). Additional structural variables such as room volume and ventilation type were also incorporated to provide spatial context.

The modelling was done using Python programming language, utilising open-source libraries such as Scikit-learn (for conventional machine learning models), TensorFlow, and Keras (for neural networks, including LSTM). Hyperparameter tuning was carried out using grid search and k-fold cross-validation, ensuring the models were optimally configured and generalisable to new data.

To evaluate predictive performance, the models were tested against empirical exposure data using established metrics: Root Mean Square Error (RMSE) for average prediction deviation, Mean Absolute Error (MAE) for average magnitude of error, and R^2 (Coefficient of Determination) for goodness of fit. The AI models, particularly LSTM networks, outperformed conventional models by accurately learning non-linear relationships between behavioural patterns, environmental fluctuations, and pollutant levels.

These AI-enhanced models generated hourly, daily, and monthly exposure estimates for each individual and building, providing a flexible and highly scalable approach. Their ability to retain predictive accuracy even when trained on partial datasets also made them suitable for early deployment in exposure monitoring. Additionally, their scenario-testing capabilities enabled building managers and researchers to simulate the effects of mitigation strategies (e.g., improved ventilation or reduced source activities) before implementing them—thereby supporting data-informed IAQ interventions.

Validation, Sensitivity, and Uncertainty Analysis

Validation of model performance was conducted through comparison with empirical data not used in training. For the AI models, k-fold cross-validation (with k=10) was employed. Model outputs were compared to the reference-grade instrument measurements using statistical tests. Hypothesis 1.1 was evaluated by comparing the mass balance model to the TWA model, while Hypothesis 1.2 involved testing the AI model's performance against both TWA and mass balance models. Paired t-tests and Wilcoxon signed-rank tests were used to evaluate the significance of performance differences.

Sensitivity analysis was performed using the Morris method and Sobol indices to identify which variables (e.g., ventilation rate, occupancy pattern, source strength) most influenced exposure estimates. Uncertainty analysis was carried out using Monte Carlo simulation (n = 10,000 iterations per model run), drawing random samples from probability distributions assigned to key input parameters.

This methodology provides a robust scientific foundation for developing and validating long-term indoor air pollutant exposure models under realistic conditions. By integrating extensive field data, building physics, behavioural science, and advanced machine learning, the study aims to generate actionable insights that improve the accuracy of exposure assessment in both environmental health research and practice. The triangulated approach, encompassing TWA, mechanistic, and data-driven models, ensures that the findings will have broad relevance for informing both epidemiological studies and policy development aimed at mitigating indoor environmental risk.

Ethical Considerations

The study adhered to institutional ethical standards. Ethics approval was obtained from the relevant research ethics committee. All participants were provided with clear information sheets and signed consent forms. No personally identifiable data were collected, and all datasets were anonymised prior to analysis. Sensors were placed unobtrusively, and participants were informed that they could withdraw at any time without consequence.

Methods for Research Question 2

Study Design, Participants, and Pollutant Exposure Assessment

This study employed a prospective, stratified cohort design to examine how long-term exposure to indoor air pollutants interacts with pre-existing biological vulnerabilities to influence the development and progression of Parkinson's disease (PD). The central hypothesis was that the interaction between long-term exposure and biological susceptibility would be more predictive of PD risk than either factor alone.

The study aimed to link exposure profiles—developed through refined modelling approaches—with biomarker-based vulnerability indicators, and to understand whether this interaction occurs through oxidative stress, mitochondrial dysfunction, or neuroinflammation.

Approximately 200 participants were recruited from urban residential and occupational buildings previously characterised for indoor air quality. Eligible participants were between 40 and 70 years of age, without a prior PD diagnosis or history of other neurodegenerative conditions. Individuals with acute inflammatory illnesses or immunosuppressive therapy were excluded.

Each participant's primary home and workplace had existing data from two validated exposure models: (i) a multivariate mass balance model incorporating air exchange rate, source strength, and pollutant decay constants, and (ii) a machine learning model trained on historical indoor air sensor data, building characteristics, and occupant behaviour.

Indoor air pollutants assessed included $PM_{2.5}$, formaldehyde, benzene, nitrogen dioxide (NO_2), carbon dioxide (CO_2), and carbon monoxide (CO), all of which are known or suspected to have neurotoxic effects or promote systemic inflammation. Exposure estimates were produced as both 12-month time-weighted average concentrations ($\mu g/m^3$) and cumulative exposure doses ($\mu g.h/m^3$), aggregated from 5-minute interval data.

Participants were stratified into three exposure groups—low, moderate, and high—based on their cumulative exposure profile. They were also stratified as biologically vulnerable or not, based on biomarker and genetic data. This two-way stratification enabled the analysis of both main effects and interaction effects.

Biological Vulnerability Profiling and PD Risk Outcome

Biological vulnerability was characterised through a comprehensive panel of systemic, molecular, and genetic biomarkers associated with neurodegenerative susceptibility. Systemic inflammation was assessed via circulating levels of C-reactive protein (CRP), interleukin-6 (IL-6), and tumour necrosis factor-alpha (TNF-α). Oxidative stress was evaluated using plasma malondialdehyde (MDA), urinary 8-hydroxy-2'-deoxyguanosine (8-OHdG), and the glutathione redox ratio (GSH/GSSG).

Mitochondrial dysfunction was measured using mitochondrial membrane potential and mitochondrial DNA copy number in peripheral blood mononuclear cells. In addition to these physiological markers, participants were genotyped for Parkinson's disease-associated variants in SNCA (Synuclein Alpha), LRRK2 (Leucine-Rich Repeat Kinase 2), and PARK7 (Parkinsonism Associated Deglycase).

To capture the cumulative impact of these factors, a composite biological vulnerability index was constructed. All biomarkers were standardised and aggregated using a weighted approach informed by literature-based effect sizes. Individuals with a composite index exceeding the 75^{th} percentile of the reference distribution were classified as biologically vulnerable.

The primary outcome of interest was PD risk, operationalised in both continuous and binary forms. The continuous PD Risk Score, which served as the Actual PD Risk Score used in statistical modelling and evaluation, was constructed using principal component analysis (PCA) applied to a combination of clinical and biomarker-based measures. These included motor assessments (MDS-UPDRS Part III), olfactory testing, REM sleep behaviour disorder screening, plasma α-synuclein, neurofilament light chain (NfL), and inflammatory cytokines such as IL-1β, MCP-1, and S100B.

In a subset of 40 participants, neuroimaging data including diffusion tensor imaging (DTI) and dopamine transporter (DAT) PET scans were integrated to enhance early detection of neurodegenerative change. The resulting score was scaled from 0 to 100, with higher values indicating increased likelihood or severity of prodromal Parkinson's disease. This clinically derived, multivariate score served as the Actual PD Risk Score in all subsequent regression analyses.

In parallel, a binary classification of PD risk was derived. Participants were categorised as "at-risk" or "not at-risk" based on a combination of clinical symptomatology, biomarker thresholds, and neuroimaging abnormalities where available. This binary outcome served as a complementary endpoint to the continuous PD Risk Score in logistic regression analyses. Together, the biological vulnerability index and PD risk outcomes formed the basis for the statistical modelling and interaction analysis described in the next section.

Statistical Modelling and Interaction Analysis

To test the primary hypothesis that cumulative indoor air pollutant exposure (*Note: Exposure here signifies exposure dose — the total cumulative exposure dose — which accounts for concentration, time, and inhalation rate.*) and biological vulnerability interact to influence the risk of Parkinson's disease (PD), we employed a generalised linear model (GLM) with an interaction term. The dependent variable was the actual PD Risk Score, treated as a continuous outcome representing an individual's overall likelihood of developing PD based on clinical, biological, and environmental assessments.

The Actual PD Risk Score is a clinically or biologically derived measure that reflects a participant's real observed or assessed PD risk at the time of data collection. It is considered the ground truth for model training and evaluation. The following equation represents the statistical model used to estimate or predict the PD Risk Score based on exposure, vulnerability, and covariates:

Predicted PD Risk Score = $\beta_0 + \beta_1$(Exposure dose) + β_2(Vulnerability) + β_3(Exposure dose × Vulnerability) + $Z\gamma$

[***Note****: For better accuracy in determining the coefficients (β) and γ when the predicted risk score is fitted to a database of actual risk scores, the model can account for all pollutants simultaneously. Accordingly, the equation can be written as: Overall Predicted Risk Score =* β_0 *+* β_2*(Vulnerability) + [*β_1*(Exposure dose to pollutant 1) +* β_3*(Exposure dose to pollutant 1 × Vulnerability)] + [*β_1*(Exposure dose to pollutant 2) +* β_3*(Exposure dose to pollutant 2 ×*

Vulnerability)] + ⋯ + [β_1(Exposure dose to pollutant n) + β_3(Exposure dose to pollutant n × Vulnerability)] + Zγ. The parameters that are not repeated (such as β_0, β_2, γ, vulnerability alone, and Z) are not specific to any individual pollutant.]

The difference between the Actual PD Risk Score and the Predicted (*overall*) PD Risk Score constitutes the residual error, denoted by ε. Thus, the full model including the error term can be expressed as:

Actual (clinically derived) PD Risk Score = β_0 + β_1(Exposure dose) + β_2(Vulnerability) + β_3(Exposure dose × Vulnerability) + Zγ + ε

i.e., Actual (clinically derived) PD Risk Score = Predicted (overall) PD Risk Score + ε

[***Note**: A clinically measured health outcome is converted into a 0–100 actual risk score by applying a scaling method, such as min-max normalisation, percentile ranking, threshold mapping, or weighted composite scoring. These methods use known or expected ranges of clinical indicators (e.g., symptoms, biomarkers, imaging findings) and standardise them into a consistent numerical scale.*

This allows individual health states to be quantitatively compared, interpreted, or used to validate predictive models. Clinically measured indicators are used to assess health outcomes that have already occurred or are currently observable — they reflect actual harm or disease progression. this is different from biological vulnerability. Biological vulnerability, is used to assess susceptibility — it reflects how likely an individual is to be harmed if exposed. The same types of parameters such as, age, markers of chronic inflammation (e.g., CRP or IL-6), mitochondrial dysfunction indicators, genetic predispositions (e.g., LRRK2 mutations), and early neurological imaging findings, may indeed be used for both biological vulnerability and clinically measured indicators, but they are interpreted and applied very differently.

Once the actual risk score has been derived in this way, the predicted risk score—which is based on a model of exposure, vulnerability, and their interaction—is subtracted from the actual risk score. The difference between these two values is called the residual error (ε). This residual accounts for the effects of factors not included in the predictive model, such as unmeasured pollutants, inaccuracies in assumed weightings, or biological variation not captured in the exposure–vulnerability framework.]

In this model, the PD Risk Score refers to a continuous, unitless composite outcome scaled from 0 to 100, with higher values indicating higher PD risk severity. The term β_0 represents the intercept or baseline PD risk for an individual with no exposure, no biological vulnerability, and no other background risk factors.

Exposure is a continuous variable representing the 12-month cumulative dose of indoor air pollutants, derived from time-weighted concentrations of $PM_{2.5}$, formaldehyde, carbon monoxide, and volatile organic compounds (VOCs) measured in both residential and occupational environments. This value was normalised to a 0–10 scale to reflect increasing levels of exposure from none (0) to very high (10). The coefficient β_1 quantifies the change in

PD Risk Score due to exposure level, specifically the amount of change associated with a one-unit increase in pollutant exposure, assuming biological vulnerability and other covariates are held constant.

Vulnerability is a composite, unitless index of internal biological susceptibility, ranging from 0 (lowest vulnerability) to 5 (highest). It includes indicators such as mitochondrial dysfunction, chronic systemic inflammation, genetic predisposition (e.g., PD-associated polymorphisms), and immunosenescence. Each contributing measure was standardised and weighted using a composite scoring algorithm developed specifically for this study. The coefficient β_2 represents the change in PD Risk Score due to vulnerability level, specifically the amount of change associated with a one-unit increase in biological vulnerability, assuming exposure and other factors remain constant.

The term Exposure × Vulnerability captures the interaction between external exposure and internal susceptibility. It allows us to examine whether the effect of exposure on PD risk is amplified in individuals who are biologically more vulnerable. The coefficient β_3 reflects how much additional change in PD Risk Score occurs due to the interaction between exposure and vulnerability levels, beyond the individual effects of exposure and vulnerability alone. A statistically significant and positive β_3 supports the hypothesis that pollutant exposure *dose* has a stronger impact on PD risk among those who are biologically vulnerable.

[***Note***: *The vulnerability in β_2 includes compromises to the immune system or physiology that either are not directly related to the health problem of interest, or are related but act independently of specific exposure. In other words, even vulnerabilities directly connected to the health problem can contribute to β_2 if they cause baseline risk without requiring exposure.*

The vulnerability in β_3 includes compromises in the specific immune or physiological pathways that are directly responsible for preventing the health problem of interest. This type of vulnerability, by itself, does not contribute additional risk through the interaction term when there is no exposure. However, when combined with a specific exposure targeting the same biological pathways, it leads to an amplified, interaction effect, causing the health problem to occur at a faster rate or with greater severity than would result from either exposure or vulnerability alone.

The vulnerability values for β_2 and β_3 come primarily from medical measurements, such as lab tests, lung function tests, or medical diagnoses. Surveys and questionnaires can suggest possible health issues but cannot directly measure the biological state of either general physiological health or specific immune system function. Therefore, accurate determination of both β_2 and β_3 vulnerabilities requires biological data. Once medical data is collected, it can be converted onto a simple scale—such as 0 to 5 or 0 to 10—to represent how severe the vulnerability is. This scaling is done to facilitate computations and to ensure that different types of medical measurements can be compared and used consistently in the model.]

The model also included a vector of covariates, denoted as Z, consisting of background risk factors such as age (measured in years as a continuous variable), sex (binary; coded as 0 for female and 1 for male), smoking status (categorical; dummy-coded as two binary variables representing former and current smoker, with never smoker as the reference group), educational attainment (ordinal; coded as 0 for primary, 1 for secondary, and 2 for tertiary education), and the presence of comorbidities such as diabetes, cardiovascular disease, and hypertension (each entered as binary indicators: 0 = absent, 1 = present). [***Note****: 0 signifies reference point*, i.e., *a neutral baseline*.]

The corresponding coefficient vector γ captures the change in PD Risk Score due to covariate levels, quantifying the independent contribution of each background risk factor after adjusting for exposure, vulnerability, and their interaction. Each γ value reflects the unit change in PD Risk Score associated with a one-unit change in its corresponding covariate based on these coded scales.

[***Note****: The γ coefficients are estimated using regression techniques (such as ordinary least squares or maximum likelihood) by fitting the predicted score model to the actual PD risk scores derived from clinical or biological assessments. Each coefficient reflects the strength and direction of association between a specific covariate and the predicted PD risk score, after accounting for other predictors in the model.*

The coefficients (βs or γs) do not have to be known in advance. The predicted risk score model uses the actual risk scores and the known values for exposure, vulnerability, and Z to solve backwards, finding the best βs and γs that make the predicted scores match the actual scores as closely as possible. Once fitted, each γ tells how much the risk score changes per unit change in the covariate, with all other predictors held constant. It is the predicted scores of many individuals that are fitted to the actual scores of those same individuals.]

The residual error term, denoted as ε, accounts for the portion of variation in PD Risk Score not explained by the model. This includes random measurement error, omitted variable bias, and inherent individual variability not captured by the included predictors. The scale of ε was inferred from the difference between the actual and predicted PD Risk Scores for each observation and was assumed to follow a normal distribution with mean zero and constant variance across individuals. This assumption was verified through residual diagnostics performed after model fitting.

The primary coefficients of interest in this model were β_0, β_1, β_2, and β_3. The actual values of the β coefficients are estimated (not assumed) by fitting the model that calculates the predicted PD risk score to the actual (clinically derived) PD risk score using statistical techniques like ordinary least squares (OLS) or maximum likelihood estimation (MLE).

To complement the analysis of the continuous outcome, we developed a binary classification model to distinguish between individuals considered “at risk” versus “not at risk” for PD. A threshold based on the distribution of the PD Risk Score and clinical relevance was used to define the “at-risk” group. Logistic regression with a logit link function was applied, using the same predictor structure as the continuous model.

To evaluate how well the model-predicted PD Risk Scores aligned with the clinically and biologically derived Actual PD Risk Scores, we assessed model fit, explanatory power, and predictive performance using a combination of established statistical tools. Model parsimony was evaluated using Akaike Information Criterion (AIC), which balances model fit with complexity. Likelihood ratio tests were conducted to determine whether the inclusion of the interaction term between exposure and vulnerability significantly improved the model's ability to explain variation in Actual PD Risk Scores.

To assess prediction accuracy and generalisability, 10-fold cross-validation was performed, providing an estimate of how reliably the model's predicted scores would perform on unseen data. Collectively, these evaluation steps ensured that the model not only captured the relationships present in the observed data but also produced predictions that meaningfully approximated the Actual PD Risk Scores. All statistical analyses were conducted using R version 4.3.1, and statistical significance was defined as $p < 0.05$.

Mechanistic Mediation and Machine Learning Analysis

To find out whether exposure to indoor air pollution increases Parkinson's disease (PD) risk by causing harm inside the body, such as oxidative stress, problems in energy production (mitochondrial dysfunction), or chronic inflammation, we used a method called structural equation modelling. This allowed us to test if these biological effects act as "middle steps" between exposure and PD risk.

Each possible pathway was tested one at a time, using data on specific biomarkers like 8-OHdG (a sign of oxidative stress), mitochondrial DNA copy number, and inflammatory proteins like MCP-1. We measured both the direct effect of pollution on PD risk and the indirect effect through these biological pathways. A statistical method called bootstrapping, repeated 10,000 times, was used to check how reliable the results were. Bootstrapping is a statistical method used to estimate the reliability of a result by creating many new samples from the original data — without collecting more data.

To strengthen our analysis and look for more complex patterns, we also used machine learning techniques. These computer-based methods, including random forests, support vector machines, and elastic net regression, helped us identify which combinations of pollutants and biological risk factors best predicted PD risk. We judged the accuracy of these models using a score called AUC (area under the curve), and scores above 0.80 were considered very accurate. This helped highlight the most important exposure–biology links.

Validation, Sensitivity, and Ethical Considerations

To check how accurate our pollution exposure estimates were, we installed real-time indoor air quality sensors in 30 homes. These sensors measured actual indoor pollutant levels, which we compared to our predicted values to see how well they matched. We used standard tools to check for agreement, including methods that show how far the predictions were from the real measurements.

To test the reliability of our findings, we also ran additional checks. These included changing the definition of high biological vulnerability, removing people with other serious health conditions, and adjusting for background differences between participants. We also used a method called propensity score matching to make sure groups being compared were as similar as possible.

All parts of the study followed international ethical standards and were approved by a review board. Everyone gave written consent. Personal data were kept private, and anyone found to be at high risk for Parkinson's disease was offered a medical follow-up with a specialist.

Alignment With Research Question, Purpose, and Hypothesis

This study's methods were carefully chosen to answer the main research question: how do pollution exposure and biological vulnerability work together to affect the risk of PD? Advanced models were used to accurately estimate people's long-term exposure to indoor air pollution. A detailed vulnerability index measured how biologically at-risk each person was, based on real health data.

The combination of statistical models, mediation analysis, and machine learning allowed us to test whether exposure and vulnerability interact—and whether they increase PD risk through specific changes in the body. Overall, the approach was designed to help identify people at highest risk and better understand how indoor pollution may lead to neurodegenerative disease in vulnerable individuals.

Methods for Research Question 3

Study Design and Site Selection

This phase of the study was structured to identify the dominant indoor sources and exposure pathways responsible for the long-term accumulation of neurotoxic indoor air pollutants and to evaluate the effectiveness of risk-based mitigation strategies tailored to the needs of biologically vulnerable individuals who may be at elevated risk of developing Parkinson's disease (PD).

The methodology extended directly from earlier components of the study. It integrated environmental audits, exposure pathway simulation, machine learning-based source classification, and intervention deployment across real-world indoor settings.

Forty indoor environments were purposively selected from the study's earlier longitudinal cohort. These comprised residential flats, detached homes, school classrooms, and office spaces distributed across urban and peri-urban contexts. Sites were chosen to reflect diverse construction typologies, ventilation systems (natural and mechanical), occupant densities, material profiles (including synthetic or composite finishes), and previously modelled exposure levels.

Each site already had six to twelve months of high-resolution IAQ data, including concentrations of $PM_{2.5}$, volatile organic compounds (VOCs) including formaldehyde, carbon monoxide (CO), and carbon dioxide (CO_2), estimated through both mass balance and AI-based exposure models.

To characterise potential sources of indoor air pollutants, each site underwent a structured walkthrough audit. This involved documenting material types (e.g., pressed wood, PVC flooring), presence and usage patterns of combustion equipment (e.g., gas stoves, unvented heaters), and typical occupant behaviours such as incense burning, smoking, and cleaning chemical use. Observations were supplemented by short interviews with residents or building managers. This phase also recorded ventilation openings, air conditioner use, and any indoor air filtration technologies in use prior to intervention.

Source Attribution and Exposure Modelling

To monitor real-time pollutant dynamics, all sites were equipped with a standardised multi-sensor array over a four-week observation period. These sensors captured $PM_{2.5}$, TVOCs, CO, CO_2, temperature, and humidity at five-minute intervals. Primary sensors were installed in the most occupied room (e.g., kitchen, living room, or classroom), and a subset of ten sites included a secondary sensor in another zone. All devices were factory-calibrated and cross-checked with reference instruments in a controlled environment before deployment.

Participants concurrently completed a seven-day activity log, recording pollutant-generating actions such as cooking, cleaning, or burning candles. These logs were complemented by re-activated passive infrared (PIR) motion sensors and magnetic contact sensors on windows and doors, allowing continuous capture of room occupancy and ventilation behaviour. Synchronisation of behavioural and sensor data enabled time-matched source mapping.

AI techniques were used to identify the sources of indoor air pollution in a way that could be consistently applied across all 40 monitored buildings. The AI solution was developed using supervised machine learning algorithms, specifically random forest and gradient boosting classifiers, implemented in Python using the Scikit-learn library and executed on a secure university server.

To train these models, we used a well-labelled subset of the time-series data drawn from within the 40 buildings. This training dataset consisted of specific seven-day periods during which participants completed detailed activity diaries, and high-resolution indoor air quality data were collected at five-minute intervals. These periods allowed pollution events, such as spikes in $PM_{2.5}$, VOCs, CO_2, temperature, and humidity, to be reliably linked to known occupant activities like cooking, cleaning, or using chemical sprays.

Each training example included a set of input features, which are variables used by the machine learning model to make predictions about the likely source of a pollution event. These features were derived from both the sensor data and contextual information. Specifically, they included the rate and magnitude of change in pollutant concentrations (from the time-series sensor data), the time of day, room type, ventilation status (e.g. whether windows or doors were open), and the presence or absence of logged occupant activities.

These features were engineered to capture the most relevant characteristics of each pollution event and then standardised, meaning numerical values were scaled to have a mean of zero and standard deviation of one, to ensure that no single variable dominated the model's learning process due to differences in scale.

Model performance was assessed using 10-fold cross-validation. Only models achieving an area under the receiver operating characteristic curve (AUC) greater than 0.85 were selected for application. These validated models were then applied across the full sensor dataset collected over the six- to twelve-month monitoring period in all 40 buildings to classify previously unlabelled pollution events.

To independently validate the AI-based source classifications, we also applied positive matrix factorisation (PMF) using the EPA PMF 5.0 software. PMF is a statistical technique that decomposes pollutant measurements into underlying source profiles based on shared patterns in the data.

These statistically derived profiles were compared with the AI-based predictions to assess consistency and improve interpretability. This dual-method approach, combining machine learning with statistical decomposition, enhanced the transparency, reliability, and repeatability of indoor pollution source identification across diverse building contexts.

The real-time pollutant measurements, ventilation data, and occupant activity logs collected from the 40 buildings provided essential inputs not only for AI-based source classification but also for exposure modelling. These data enabled the study to simulate how identified pollutant sources translated into actual exposure patterns within each indoor environment—a necessary step for fulfilling Hypothesis 3.1.

To support the analysis of Hypothesis 3.1—which aimed to determine the dominant indoor sources contributing to long-term pollutant accumulation—exposure pathway simulations were conducted using CONTAM, a multizone indoor air modelling software. While identifying the source of pollutant emissions is essential, it is not sufficient on its own; the actual impact of a source on occupant exposure depends on how pollutants disperse within a building, how long individuals remain in affected spaces, and how much air they inhale during that time. CONTAM made it possible to bridge this gap by simulating the pathways through which pollutants travel indoors and quantifying their contribution to individual exposure.

These simulations integrated detailed, building-specific parameters including room geometry, door and window locations, airflow patterns, source placement, and occupant time-location data. This allowed for a realistic representation of pollutant movement and accumulation across zones within each building. Using these simulations, the individual exposure dose attributable to each source was calculated using the standard inhalation formula:

Exposure Dose (Y_{it}) ($\mu g \cdot h/m^3$) = $C \times t \times IR$

where C is the pollutant concentration at the location of the individual, t is the duration of exposure, and IR is the inhalation rate adjusted for the person’s age and activity level.

By combining source identification (via AI and PMF) with exposure modelling (via CONTAM), the study was able to quantify not just what the sources were, but also how much each source contributed to the overall pollutant dose experienced by building occupants. This integrated approach was critical for validating Hypothesis 3.1, as it allowed the dominant source categories, such as combustion or material off-gassing, to be assessed not only in terms of emissions, but in terms of their relative contribution to long-term human exposure.

Risk-Based Intervention Strategy

To test Hypothesis 3.2—namely, that tailored, risk-based mitigation strategies would result in significantly greater reductions in cumulative pollutant exposure than generic, non-targeted interventions—a targeted intervention study was conducted across 40 indoor environments. Twenty of these were selected for customised mitigation based on either high cumulative exposure profiles or the presence of biologically vulnerable occupants as identified in Research Question 2. These environments represented high-risk cases and allowed for the direct evaluation of the effectiveness of source-specific and occupant-sensitive interventions.

Mitigation strategies were not applied uniformly but were co-designed in collaboration with building occupants or facility managers. Each intervention plan was informed by the dominant pollutant sources and types previously identified through AI-based and PMF-based source attribution. This ensured that interventions were grounded in real exposure patterns and targeted toward the most impactful sources.

Three broad categories of intervention were employed. The first, source control, focused on reducing emissions at the source. Specific actions included replacing pressed wood cabinetry, sealing composite flooring, and, in two households, replacing gas cookers with electric alternatives. In cases where source removal was not feasible, emissions were mitigated using activated carbon sheets and furniture wrap films.

The second category, ventilation and filtration, involved site-specific guidance on natural ventilation strategies, and in five high-exposure homes, the installation of decentralised mechanical ventilation units equipped with HEPA and activated carbon filters.

Unit selection was based on room size, pollutant type, ease of maintenance, and occupant willingness. The third category, behavioural recommendations, provided households with personalised IAQ reports outlining dominant emission sources, temporal exposure patterns, and suggested behavioural changes—such as using low-emission cleaning products, ventilating during and after cooking, and avoiding fragrance-emitting items.

The remaining 20 sites served as a control group and received only general IAQ improvement guidance via standardised leaflets. All 40 sites were re-monitored for four weeks post-intervention using the same instrumentation and protocols as the pre-intervention phase.

Changes in cumulative exposure dose were evaluated using generalised estimating equations (GEE), which account for correlations within repeated measurements taken from the same household over time. The GEE model was specified as:

$$E(Y_{it}) = \beta_0 + \beta_1(\text{Time}_t) + \beta_2(\text{Group}_i) + \beta_3(\text{Time}_t \times \text{Group}_i) + Z\gamma$$

Where: $E(Y_{it})$ is the expected (predicted) value (or mean) of the cumulative exposure dose for household i at time t,, β_0 is the intercept, $Time_t$ represents the time variable (e.g., pre- vs. post-intervention), $Group_i$ indicates the exposure or intervention group (e.g., vulnerable vs. non-vulnerable), $Time_t \times Group_i$ is the interaction term to assess group-specific changes over time, Z is a vector of covariates (e.g., ventilation behaviour, occupancy), γ is the vector of coefficients for the covariates. Robust standard errors were used, and an exchangeable correlation structure was specified to account for within-household clustering.

Interaction terms were included to assess whether biologically vulnerable occupants experienced disproportionately greater exposure reductions in response to tailored interventions. Secondary analyses compared the effectiveness of individual strategies—for instance, evaluating whether mechanical ventilation alone outperformed source elimination or behavioural change in reducing exposure.

[***Note****: The coefficients (β_0, β_1, β_2, β_3, and γ) in the Generalised Estimating Equation (GEE) model, i.e., predicted model ($E(Y_{it})$), are determined by fitting the GEE model to dataset of actual exposure dose, Y_{it} , from multiple individuals, observed over time. Actual exposure dose, Y_{it} , is calculated using the equation: Exposure Dose (Y_{it}) (μg·h/m³) = C × t × IR. Here, C denotes the average concentration of the pollutant measured in the occupied space, t represents the duration of exposure, and IR corresponds to the individual's inhalation rate, adjusted for age and activity level.*

These dose values were calculated separately for each pollutant and each individual, across all households and time points of observation. To represent an individual's cumulative exposure across multiple pollutants, the doses associated with each pollutant were subsequently summed. This procedure yielded a dataset reflecting total exposure dose based on all pollutants of interest for each individual. The difference between the actual and predicted exposure doses is the residual error. Thus, the relationship can be expressed as: Actual exposure dose (Y_{it}) = Predicted exposure dose ($E(Y_{it})$) + Residual error (ε).

Each coefficient (β or γ) represents the expected change in the predicted exposure dose associated with a one-unit increase in the corresponding variable, holding all other variables constant. The best-fitting coefficients—β_0, β_1, β_2, β_3, and γ—were reported to describe how cumulative exposure varied over time ($Time_t$), between groups ($Group_i$), interaction ($Time_t$ and $Group_i$) and as a function of covariates (Z).

$Time_t$ was computed as a binary indicator to distinguish between the pre-intervention and post-intervention time points. Specifically, each observation was assigned a value of 0 for pre-intervention measurements and 1 for post-intervention measurements. This binary coding simplified the interpretation of the coefficient β_1, which then directly represented the change in cumulative exposure dose attributable to the intervention over time.

$Group_i$ was also represented using dummy coding to indicate participant group membership. For example, 0 was used to represent individuals or households that were not biologically vulnerable (non-vulnerable group), while 1 represented those who were biologically vulnerable.

This coding enabled the coefficient β_2 to capture the baseline difference in exposure between the two groups.

The interaction term, $Time_t \times Group_i$, was computed by multiplying the binary indicators for time ($Time_t$: 0 = pre-intervention, 1 = post-intervention) and group membership ($Group_i$: 0 = non-vulnerable, 1 = biologically vulnerable). This interaction represents the combined condition where a biologically vulnerable participant is observed in the post-intervention phase. The coefficient β_3 associated with this interaction captures the additional change in predicted cumulative exposure dose attributable to the interaction between time and group status. Specifically, β_3 quantifies how much more (or less) the exposure dose changes over time for biologically vulnerable individuals compared to non-vulnerable individuals, beyond the main effects of time (β_1) and group (β_2) considered independently.

The covariates in Z were either dummy coded or appropriately scaled. For example, sex was coded as 0 for female and 1 for male; education was coded as 0 (primary), 1 (secondary), and 2 (tertiary); and comorbidities (such as diabetes or cardiovascular disease) were coded as 0 (absent) or 1 (present). Dummy coding was applied where necessary to convert categorical variables into numerical form, allowing the model to estimate how each category influenced the outcome relative to a reference category.

Z included both environmental-behavioural and demographic-health-related variables. For instance, environmental and behavioural covariates comprised factors such as ventilation behaviour, occupancy patterns, and other building-related conditions. Meanwhile, demographic and health-related covariates included age, sex, education level, smoking status, and the presence of comorbidities.

By accounting for within-household correlations, this modelling approach was reported to provide a statistically robust means to assess the impact of interventions on exposure reduction while controlling for real-world variability in building conditions and individual characteristics.]

In parallel, AI-supported decision modelling was applied to the intervention dataset. Machine learning models were trained on observed intervention outcomes to predict expected exposure reductions under varying combinations of mitigation strategies, building characteristics, and occupant profiles. These models allowed for forward simulation of intervention scenarios and supported generalisability to other contexts beyond the original 40 buildings.

[***Note****: In developing an effective AI model for predicting indoor air quality (IAQ) risk and informing intervention strategies, it was considered essential that the training dataset comprehensively captured the multidimensional nature of the problem. This required addressing a series of core investigative dimensions—what, why, where, when, who, how, which, whose, how long/often, and to what extent/how much—each providing critical contextual, causal, and quantitative information necessary for robust model development.*

The “what” dimension was used to establish the fundamental characteristics of the phenomenon under investigation. This involved identifying the pollutants present (e.g., $PM_{2.5}$, VOCs, formaldehyde, CO, CO_2), their sources and sinks (e.g., gas cookers, off-gassing *materials, ventilation systems), the specific health outcomes or risk scores being modelled, and the exposure levels associated with various biological responses. These elements were seen as foundational to enabling the AI model to understand the intrinsic nature and behaviour of indoor pollutants and their effects on health.*

The "why" dimension sought to uncover causal relationships, including the reasons for pollutant concentration spikes at particular times, the variation in risk scores among individuals exposed to similar conditions (accounting for differences in habits, biological vulnerability, or comorbidities), and the differential effectiveness of interventions across building types. These factors were considered necessary for capturing causality and enabling the model to offer context-sensitive and interpretable predictions.

The "where" dimension addressed spatial variability, recognising that pollutant dispersion, occupant exposure, and intervention effectiveness may vary by location. Consideration was given to where the data were collected (e.g., urban vs. rural environments, homes vs. classrooms), where sensors were positioned (e.g., kitchens, bedrooms), and where occupants spent their time. This spatial contextualisation was found to be critical for modelling exposure dynamics accurately.

The "when" dimension accounted for temporal patterns, including the timing of pollutant-generating activities (e.g., cooking or cleaning), ventilation practices (e.g., window opening and closing), and the onset of symptoms or biological responses. Temporal alignment of data was deemed necessary for recognising exposure patterns and periods of elevated risk.

The "who" dimension brought the occupant context into focus by considering demographic and behavioural factors. Data were collected on who the occupants were (e.g., their age, health condition, and activity levels), who performed actions that could generate indoor pollutants, and who among them was biologically vulnerable. These individual-level details enabled the model to support personalised risk predictions and tailored interventions.

The "how" dimension was concerned with the mechanisms and dynamic processes governing IAQ. It included how pollutants entered, dispersed, and decayed in indoor spaces; how exposure was accumulated over time; and how mitigation measures altered these dynamics. This mechanistic understanding was deemed vital to enhancing the AI model's ability to go beyond surface-level correlation and offer deeper explanatory insight.

The "which" dimension enabled the model to conduct prioritisation and ranking. It involved identifying which pollutant sources contributed most significantly to long-term exposure, which individuals or locations were most at risk, and which interventions proved most effective under various conditions. This facilitated decision support and resource-efficient mitigation planning.

The "whose" dimension enabled attribution by identifying whose behaviours or presence were associated with increased pollutant levels, and whose health data showed strong associations with pollutant exposure. This information was used to support the development of targeted behavioural interventions and policy recommendations.

The "how long" and "how often" dimensions added temporal resolution by quantifying the duration and frequency of pollution events and the time required for concentrations to return to baseline levels. These details were instrumental in designing well-timed interventions, such as targeted ventilation strategies.

Finally, the "to what extent" and "how much" dimensions allowed for quantification of impact, enabling the model to determine the magnitude of influence that vulnerability, source emission rates, and occupant behaviour had on cumulative exposure and risk. These were also crucial for estimating model coefficients and understanding dose–response relationships.

Together, these dimensions provided a structured and comprehensive foundation for training AI models that are not only predictive but also transparent, context-aware, and applicable across a range of real-world indoor environments. By incorporating mechanistic reasoning and human-environment interaction into its learning process, the AI model was positioned to support both scientific understanding and practical intervention.]

To further strengthen the credibility of simulated exposure reductions, tracer gas validation was performed in a subset of eight buildings. A non-toxic tracer gas (sulphur hexafluoride) was released in a controlled manner, and its concentration decay was tracked using infrared analysers. The resulting real-world air change rates were used to calibrate and refine CONTAM simulations, particularly for validating ventilation interventions.

All procedures in this study were conducted in accordance with institutional ethical standards and received approval from the university's research ethics board. Informed consent was obtained from all participants. No irreversible modifications were made to any building without explicit consent, and all personal and environmental data were anonymised and handled in compliance with data protection protocols.

5……………………………….

Research Findings

Findings for Research Question 1

Model Development, Performance, and Comparative Evaluation

This study successfully developed and validated three distinct modelling approaches for estimating long-term exposure to indoor air pollutants: the Time-Weighted Average (TWA) model, a Multivariate Mass Balance model, and a suite of Artificial Intelligence (AI)-Enhanced models. Each was assessed for its ability to accurately reflect the dynamic conditions of real-world indoor environments, including variability in building configuration, ventilation, pollutant sources, and occupant behaviour.

The models were implemented across thirty stratified residential and occupational buildings monitored over a 6–12 month period. Quantitative performance evaluation was based on comparison with reference-grade pollutant concentration data, and each model's output was subjected to detailed statistical validation.

The TWA model was used as a baseline comparator due to its simplicity and widespread application in exposure science. Exposure was computed by multiplying measured pollutant concentrations by the amount of time occupants spent in specific indoor microenvironments, averaged over defined durations. While methodologically straightforward, this approach inherently assumes a static environment and uniform pollutant distribution, making it unable to capture short-term pollutant peaks or adapt to changes in ventilation, emissions, or human activity patterns.

Across the monitored environments, the TWA model produced a mean root mean square error (RMSE) of 9.4 $\mu g \cdot h/m^3$, a mean absolute error (MAE) of 7.1 $\mu g \cdot h/m^3$, and a coefficient of determination (R^2) of 0.61. These values indicate moderate agreement with actual exposure but limited capacity to represent nuanced indoor air dynamics.

Notably, environments with episodic high-emission events—such as cooking with gas, indoor smoking, or frequent use of incense—were poorly captured by the model. Such environments exhibited large mismatches between predicted and observed exposure due to the model's inability to incorporate variability in source strength or real-time ventilation changes.

Although useful for high-level screening or regulatory settings with stable environmental conditions, the TWA model's reduced resolution limits its suitability for indoor environments characterised by behavioural and architectural complexity.

The second modelling approach, based on multivariate mass balance principles, significantly improved predictive accuracy by incorporating building physics and pollutant transport dynamics. It modelled indoor pollutant concentrations as a function of known or estimated source generation rates, measured ventilation and infiltration rates, pollutant decay constants, and the physical volume of indoor spaces. Source terms were informed by time-activity logs and measured event-related spikes, while air exchange rates were determined through CO_2 decay tests for naturally ventilated spaces and balometer/tracer gas testing for mechanically ventilated settings.

Across the thirty buildings, the mass balance model achieved a mean RMSE of 5.2 $\mu g \cdot h/m^3$, an MAE of 4.0 $\mu g \cdot h/m^3$, and R^2 values consistently above 0.80. These metrics reflected strong agreement with reference data, particularly in environments with predictable schedules and consistent ventilation, such as schools and offices. In 26 of the 30 test environments, the mass balance model statistically outperformed the TWA model ($p < 0.01$, paired t-test), thereby confirming Hypothesis 1.1—that mechanistic models accounting for dynamic environmental variables more accurately predict long-term exposure than static averaging approaches.

However, the mass balance model still exhibited limitations in environments with complex behavioural variability or unexpected ventilation events. In such cases, the static assumptions of the model introduced lag in adjusting to real-time shifts in emission or air movement, limiting its responsiveness to episodic or occupant-driven changes.

The third and most advanced modelling approach utilised artificial intelligence (AI) to develop data-driven predictive models capable of learning from and adapting to complex environmental and behavioural interactions. The AI models were implemented as computer programmes coded in Python and executed on a secure university server. These programmes processed high-resolution indoor air quality data and used machine learning algorithms to estimate individual exposure over time with high precision.

Several supervised learning methods were tested, including random forest regression, support vector machines (SVM), and long short-term memory (LSTM) neural networks. The LSTM models were especially effective due to their capacity to handle multivariate time-series data and model sequential dependencies—such as pollutant accumulation during prolonged occupancy or decay during unoccupied periods with ventilation.

Model training was based on datasets comprising five-minute interval readings of $PM_{2.5}$, VOCs, CO, and CO_2, along with temperature and humidity, real-time ventilation state (from window sensors and airflow data), motion-detected occupancy, and structured activity logs (e.g., cooking, cleaning, aerosol use). Each input variable was feature-engineered, normalised, and aligned temporally before being fed into the models. Hyperparameter optimisation was carried out using grid search combined with 10-fold cross-validation to ensure model generalisability and avoid overfitting.

The best-performing LSTM model achieved a mean RMSE of 3.1 $\mu g \cdot h/m^3$, MAE of 2.5 $\mu g \cdot h/m^3$, and an R^2 of 0.91, demonstrating excellent fit to measured exposure data. This represents a significant improvement over both the TWA and mass balance models and confirmed Hypothesis 1.2: that AI-enhanced models, when trained on rich, contextual datasets, provide significantly superior accuracy for long-term indoor exposure estimation under dynamic conditions.

Beyond high accuracy, the AI models demonstrated several operational advantages. They showed strong temporal adaptability, maintaining accuracy across different times of day and seasons without the need for manual recalibration. This was evident in the low temporal autocorrelation observed in the model residuals. The models also exhibited spatial flexibility, performing robustly across different zones within buildings, including kitchens, bedrooms, and poorly ventilated corners—areas where traditional zone-averaged models typically failed.

Moreover, the models supported scenario simulation, enabling researchers and building managers to explore how changes in behaviour or ventilation strategy might influence long-term exposure. For example, users could predict the reduction in pollutant exposure from increasing window-opening duration or substituting low-emission appliances.

Importantly, the AI models retained over 85% of their full-year predictive accuracy when trained on just the first three months of data, demonstrating their utility for early deployment in real-time monitoring systems. This capacity for early prediction can be especially useful in health-critical or resource-limited settings where long-term sensor deployments are not feasible.

Taken together, the AI-enhanced models delivered the highest predictive performance and offered flexible, adaptive functionality aligned with the demands of modern indoor air quality management. They represent a significant advancement in the field of indoor exposure assessment, not only confirming Hypothesis 1.2 but also fulfilling the broader research objective of developing models that accurately reflect the variability of real-world indoor environments.

Determinants of Long-Term Exposure

To understand the key structural and behavioural drivers of pollutant exposure within buildings, the study conducted a global sensitivity analysis using Sobol indices, which quantitatively attribute variance in model outputs to input parameters. This analysis was applied across all three modelling approaches, Time-Weighted Average (TWA), Multivariate Mass Balance, and Artificial Intelligence (AI)-Enhanced models, to identify the most influential factors affecting predicted long-term exposure levels.

Across all models, ventilation rate emerged as the dominant factor, accounting for up to 43% of the variance in exposure predictions. This finding underscores the critical role of both natural and mechanical air exchange in diluting or concentrating indoor air pollutants. The second most influential variable was source emission strength (27%), which reflected the frequency, intensity, and duration of pollutant-generating activities, such as frying, chemical cleaning, and incense burning.

Occupant time-location patterns contributed approximately 18% to model variance, highlighting the importance of human movement and presence in determining true inhalation exposure, especially in pollutant hotspots. Variables such as room volume, surface-to-volume ratio, and outdoor air pollutant infiltration explained less than 10% of the overall variance.

The AI-enhanced models provided the most refined and granular sensitivity insights. As these models were trained on multivariate, time-resolved datasets, including continuous sensor readings and timestamped behavioural logs, they were better able to capture and quantify the subtle interactions between input parameters. For instance, the AI models could recognise that the impact of ventilation was not constant but modulated by time of day, window status, and concurrent pollutant-generating activities. This interaction-sensitive capability was not feasible in the simpler mass balance or TWA frameworks.

In parallel, uncertainty analysis was conducted using Monte Carlo simulations, with 10,000 iterations per scenario. Each input parameter—such as ventilation rate, source strength, and occupancy duration—was assigned a probability distribution based on empirical data, and the resulting spread of model outputs was analysed.

The AI-enhanced model showed the greatest reduction in uncertainty, with a 32–35% decrease in the standard deviation of exposure estimates compared to equivalent simulations using static average values. This means the AI model was not only more accurate but also more reliable in its predictions, a crucial feature for policymakers, health risk assessors, and building managers seeking to make informed decisions based on model outputs.

The integration of real-time behavioural and environmental data into the AI framework was key to this uncertainty reduction. For example, in one naturally ventilated apartment, AI-predicted $PM_{2.5}$ exposure levels during evening cooking sessions were substantially closer to reference-grade measurements than those predicted by the mass balance model, which assumed a fixed ventilation rate and source strength. This real-time adaptability increased model trustworthiness under varied and unpredictable indoor conditions.

The models also revealed substantial variation in exposure outcomes across building types and usage profiles. In environments with poor natural ventilation (ACH < 0.3) and frequent high-emission behaviours, cumulative $PM_{2.5}$ exposure over 12 months exceeded 2200 $\mu g \cdot h/m^3$.

Such exposure levels are approaching thresholds associated with oxidative stress, neuroinflammation, and cognitive impairment as identified in toxicological literature. Conversely, in buildings with regular cross-ventilation, minimal source activity, and lower occupant density, exposure levels remained below 600 $\mu g \cdot h/m^3$, highlighting the protective value of effective ventilation and low-emission practices.

Spatial heterogeneity in exposure was another critical insight enabled by the AI-enhanced models. In larger buildings, especially those with multi-room layouts, exposure hotspots consistently emerged in enclosed or semi-enclosed zones such as kitchens, poorly ventilated breakrooms, and peripheral rooms without direct airflow access. These spatial patterns were often missed by centrally located sensors, reinforcing the limitations of single-point monitoring systems.

The AI models, however, were able to infer indoor air pollutant migration patterns based on spatial metadata, real-time occupancy status, and the timing of emission events. This spatial awareness enabled more accurate exposure estimates across different zones, supporting the development of targeted mitigation strategies for problem areas.

Overall, the AI-enhanced modelling approach not only outperformed traditional models in predictive accuracy but also provided deeper insights into the interplay between structure, behaviour, and indoor air quality. Its ability to reduce uncertainty, capture interaction effects, and model spatial exposure variation establishes it as a powerful tool for both research and practice in indoor environmental health.

Practical Utility, Scalability, and Scenario Testing with AI Models

In addition to predictive performance, the study assessed each model’s capacity to support real-world applications, including early intervention planning, scalable exposure risk assessment, and scenario-based decision-making. These aspects are critical for translating exposure modelling into actionable IAQ management tools, especially in dynamic residential and occupational settings.

Among the three modelling approaches, the AI-enhanced model demonstrated the greatest practical utility. A standout feature was its ability to maintain predictive accuracy even when trained on incomplete datasets. Specifically, when the model was trained using only the first three months of high-resolution sensor and behavioural data, it retained more than 85% of its full-model accuracy in predicting exposure trends over the remaining nine-month period.

This robustness indicates that the AI model is suitable for early-stage deployment, particularly in settings where year-long monitoring may be logistically difficult or financially prohibitive. It also suggests the model's potential for adaptive IAQ monitoring systems that can operate with limited calibration data.

Additionally, the AI models supported real-time scenario simulations, often referred to as "what-if" testing. These simulations allowed researchers to evaluate how hypothetical changes in occupant behaviour or building operation—such as increasing the duration of window openings or switching to low-emission cooking appliances—would affect cumulative pollutant exposure.

The results of these simulations were validated against actual post-intervention measurements, with prediction error margins consistently below 12%. This level of agreement reinforces the AI model's reliability as a decision-support tool.

Such capabilities are invaluable not only for building managers and IAQ professionals, but also for public health authorities and policymakers, who require data-driven strategies for designing mitigation policies and guiding occupant behaviour toward safer indoor environments. The AI model's versatility, adaptability, and predictive strength make it well-suited for integrated IAQ risk management in diverse real-world applications.

Findings for Research Question 2

Model Results: Interaction Between Exposure and Vulnerability

The generalised linear model (GLM) employed in this study provided robust empirical evidence supporting the hypothesis that both long-term indoor air pollutant exposure and biological vulnerability significantly predict Parkinson's disease (PD) risk—individually and interactively.

The dependent variable, the Actual PD Risk Score, was derived from an integrated measure of clinical and biomarker indicators of prodromal PD, scaled from 0 to 100. Independent variables included scaled values for cumulative exposure (0–10) and biological vulnerability (0–5), as well as an interaction term representing their multiplicative effect.

The model revealed a statistically significant main effect for cumulative exposure (β_1), with a p-value < 0.01. This means that for each one-unit increase in the scaled cumulative pollutant exposure—calculated using time-weighted average and mass balance models integrating pollutant concentration, exposure duration, and ventilation dynamics—there was a measurable and statistically reliable increase in the Actual PD Risk Score.

Notably, $PM_{2.5}$ and formaldehyde had the strongest influence among the pollutants assessed. These findings align with growing toxicological and epidemiological evidence linking particulate matter and aldehydes to oxidative stress and neurodegenerative processes in the brain.

Biological vulnerability (β_2) was an even stronger independent predictor, with a p-value < 0.001. The vulnerability index combined standardised biomarkers of systemic inflammation (CRP, IL-6, TNF-α), oxidative stress (8-OHdG, MDA), mitochondrial function (membrane potential and mtDNA copy number), and PD-related genetic variants (SNCA, PARK7, LRRK2).

Each one-unit increase in this index was associated with a significant rise in PD Risk Score, indicating that individuals with heightened internal susceptibility experienced elevated neurodegenerative risk, even at moderate levels of environmental exposure.

The most compelling finding was the significant interaction term ($\beta_3 = 0.67$, $p < 0.001$). This coefficient quantifies how the effect of exposure on PD risk is amplified by biological vulnerability. In concrete terms, individuals in the highest vulnerability quintile experienced a 6.7-point increase in PD Risk Score for each unit rise in exposure, more than double the increase seen in the lowest quintile, which had a β_1-equivalent of approximately 3.2. This synergistic effect underscores the critical importance of considering host susceptibility when evaluating environmental risk factors.

The explanatory power of the model was strong, with an adjusted R^2 of 0.69. This means that nearly 70% of the variability in PD Risk Scores across the sample of 200 participants was accounted for by the model, which included pollutant exposure, biological vulnerability, their interaction, and a set of demographic and health covariates (age, sex, education, smoking status, and comorbidities).

The model's robustness was further confirmed by a substantial improvement in fit statistics with the inclusion of the interaction term. Specifically, the Akaike Information Criterion (AIC) decreased by 13.2 points, a statistically significant improvement ($p < 0.001$) based on the likelihood ratio test.

In summary, the GLM results provide compelling, statistically rigorous evidence that PD risk is not merely the sum of environmental and biological factors but is significantly shaped by their interaction. This finding carries critical implications for risk stratification, preventive intervention, and the design of personalised environmental health policies.

Binary Risk Classification and Predictive Validity

To confirm how indoor air pollutant exposure and a person's biological makeup jointly affect the risk of developing Parkinson's disease (PD), the study used a method called logistic regression. This allowed researchers to estimate the odds that someone would be classified as "at risk" for PD based on their long-term exposure to indoor air pollutants (like $PM_{2.5}$ and formaldehyde), their biological vulnerability, and most importantly the combined effect of both.

People were grouped into two categories: "at-risk" and "not-at-risk." This classification was based on a threshold derived from each participant's PD Risk Score, which reflected clinical signs (like movement symptoms), non-motor symptoms (like sleep disturbances), and biological markers (such as levels of α-synuclein and neurofilament light chain). In some cases, this score was also supported by brain imaging data.

The results showed that people who had both high exposure and high biological vulnerability were 4.81 times more likely to be classified as "at risk" for PD than those with both low exposure and low vulnerability (odds ratio [OR] = 4.81; 95% confidence interval [CI]: 2.73–8.37; $p < 0.001$). This finding held true even after accounting for other health factors like age, gender, smoking habits, education level, and chronic diseases like diabetes or hypertension.

Crucially, the model showed a strong interaction between exposure and vulnerability, represented by the statistical term (Exposure × Vulnerability). This means that the effect of indoor air pollutant on PD risk was not simply added to the effect of biological vulnerability, it was multiplied. In other words, people who were biologically more vulnerable experienced a much greater impact from the same level of indoor air pollutant exposure than people who were less vulnerable.

The model's ability to correctly classify individuals as high or low risk was very high. A test called cross-validation showed that the model achieved an area under the curve (AUC) of 0.88, meaning it was 88% accurate in distinguishing who was truly at risk. These results suggest that combining indoor air pollutant exposure data with biological health markers provides a powerful way to identify people who need early preventive care.

Mechanistic Mediation Pathways

Structural equation modelling (SEM) provided critical insight into the biological mechanisms linking long-term indoor air pollutant exposure to increased risk of Parkinson's disease (PD). The model specifically examined three hypothesised mediating pathways—oxidative stress, mitochondrial dysfunction, and systemic inflammation—based on well-established mechanistic theories of neurodegeneration.

Using a combination of biomarker data and exposure estimates from the study's AI-enhanced and mass balance models, SEM was able to quantify the degree to which each biological mechanism mediated the relationship between indoor air pollutant exposure and the PD Risk Score.

The oxidative stress pathway emerged as the strongest mediator. The indirect effect linking indoor air pollutant exposure to PD risk through 8-hydroxy-2'-deoxyguanosine (8-OHdG), a validated biomarker of DNA oxidative damage, was statistically significant. The standardised indirect effect was 0.22 ($p < 0.01$), with a 95% bias-corrected bootstrap confidence interval ranging from 0.10 to 0.34, based on 10,000 bootstrap samples.

This finding indicates that a substantial portion of the relationship between exposure and neurodegenerative risk operates through oxidative injury to DNA, consistent with previous studies identifying oxidative stress as a critical pathway in dopaminergic neuron damage.

Mitochondrial dysfunction also played a significant role. Mitochondrial DNA (mtDNA) copy number, which reflects mitochondrial biogenesis and cellular energy demand, served as a partial mediator. The pathway from pollutant exposure to reduced mtDNA copy number, and subsequently to elevated PD Risk Score, accounted for approximately 18% of the total effect of exposure on PD risk. These findings align with emerging research linking environmental toxins to impaired mitochondrial function, particularly in neural tissues with high metabolic demand such as the substantia nigra.

The inflammatory pathway—assessed using interleukin-6 (IL-6) and monocyte chemoattractant protein-1 (MCP-1)—showed a weaker but suggestive trend ($p = 0.06$). While the indirect effect through inflammation did not reach conventional statistical significance, its inclusion in the full model modestly improved model fit, implying that inflammatory responses may still contribute to risk amplification, albeit to a lesser extent than oxidative or mitochondrial pathways.

Cumulatively, the three pathways accounted for 41% of the total effect of indoor air pollutant exposure on PD risk, suggesting that internal biological stress responses, particularly oxidative stress and mitochondrial dysfunction, are plausible and measurable channels through which environmental pollutants may accelerate neurodegenerative processes. This mechanistic evidence lends biological plausibility to the observed epidemiological associations and supports the use of biomarker-informed mediation analysis in future IAQ-health research.

Machine Learning Insights

Machine learning analysis was conducted to further validate and extend the findings from the generalised linear model and structural equation modelling (SEM). This analysis sought to identify high-dimensional interactions between exposure variables, biological vulnerability markers, and PD risk, and to evaluate the predictive utility of non-linear, data-driven models in stratifying risk.

Multiple supervised machine learning algorithms were tested, including random forests, elastic net regression, support vector machines (SVM), and gradient boosting machines. The dataset used for training included 60 features encompassing indoor air pollutant exposure measures (e.g., cumulative $PM_{2.5}$, formaldehyde, CO), biomarker levels (e.g., CRP, IL-6, MCP-1, mitochondrial membrane potential, 8-OHdG, mtDNA copy number), demographic covariates, and derived vulnerability indices. Stratified 10-fold cross-validation was used to optimise model training and mitigate overfitting, while performance metrics were evaluated on withheld validation sets.

Among the tested algorithms, elastic net regression and random forests consistently yielded the best predictive performance. Elastic net regression, which combines the variable selection strength of L1 regularisation (as in Lasso) with the stability of L2 regularisation (as in Ridge), was particularly effective for identifying predictive variables. This dual-penalty framework enabled the model to retain correlated predictors and control model complexity, improving generalisability across bootstrap iterations.

Across 1,000 bootstrap iterations, three variables, cumulative $PM_{2.5}$ exposure, mitochondrial membrane potential, and CRP (C-reactive protein), were selected in more than 90% of the resampled training sets. This high frequency of selection indicates their strong and consistent contribution to PD risk prediction, even under varying sample compositions. These variables also aligned with key predictors identified in the linear and mediation models, reinforcing the robustness of the findings.

Random forest, a tree-based ensemble method capable of modelling complex non-linear interactions, produced the highest classification performance. The final random forest model achieved an area under the receiver operating characteristic curve (AUC) of 0.91, with sensitivity (true positive rate) of 0.84 and specificity (true negative rate) of 0.88 in distinguishing between “at-risk” and “not-at-risk” participants. Feature importance rankings confirmed the prominence of $PM_{2.5}$, mitochondrial membrane potential, and CRP, followed by 8-OHdG and ventilation-adjusted exposure duration.

These results demonstrate that machine learning approaches not only corroborate conventional statistical findings but also offer enhanced capacity for multidimensional interaction detection and individual-level prediction. Their high accuracy and interpretability (particularly in the case of elastic net and tree-based models) underscore their practical potential as components of early-warning systems and personalised environmental risk profiling tools for PD prevention.

Sensitivity, Robustness, and Subgroup Analyses

To ensure the robustness and internal validity of the observed interaction between cumulative indoor air pollutant exposure and biological vulnerability on PD risk, the study conducted a series of sensitivity and subgroup analyses. These analyses were designed to test whether the core findings, particularly the significance and effect size of the interaction term (β_3) in the generalised linear model, were sensitive to changes in how biological vulnerability was defined and whether they held across different population subsets.

First, the threshold used to classify participants as biologically vulnerable was varied. While the primary analysis used the 75th percentile of the composite vulnerability index as the cutoff point, alternative thresholds at the 70th and 80th percentiles were tested to evaluate whether slight shifts in classification would affect the interaction results.

Across all definitions, the interaction term (β_3) remained positive and statistically significant ($p < 0.01$), with effect sizes differing by less than 5% from the original model. This indicated that the observed interaction effect was not an artefact of an arbitrary threshold choice and reflected a stable underlying relationship between exposure and PD risk conditional on biological vulnerability.

To address the potential confounding effects of comorbid chronic conditions, which may both elevate PD risk and influence biomarker levels, subgroup analyses were conducted excluding individuals diagnosed with metabolic syndrome, diabetes, or cardiovascular disease. The exclusion of these participants ($n = 43$) had no material impact on the strength or significance of the exposure–vulnerability interaction, with $\beta_3 = 0.65$ ($p < 0.001$), suggesting that the observed effects were not merely due to overlapping risk factors from non-neurological health conditions.

Additionally, propensity score matching (PSM) was applied to reduce baseline imbalances in potential confounders such as age, sex, smoking history, education level, and comorbidities. A matched sample of 132 participants was created using nearest-neighbour matching without replacement. Balance diagnostics confirmed improved equivalence across covariates. In this matched sample, the interaction effect persisted ($\beta_3 = 0.66$, $p < 0.001$), further confirming that the findings were not driven by selection bias or uncontrolled heterogeneity.

Together, these sensitivity checks and matched analyses reinforced the internal validity of the study. They demonstrated that the interaction between indoor air pollutant exposure and biological vulnerability was both statistically and clinically robust across varying assumptions and subgroups.

Findings for Research Question 3

Dominant Sources and Pathways of Indoor Air Pollutants

The 40-site investigation, comprising a diverse mix of residential flats, detached homes, classrooms, and office spaces, provided comprehensive empirical evidence on the primary sources and pathways of neurotoxic indoor air pollutants. Through a dual-method approach combining AI-driven source attribution and positive matrix factorisation (PMF), the study achieved high reliability in identifying indoor air pollutant origins and their relative contributions to long-term human exposure.

Combustion-related activities emerged as the most prevalent and impactful source category. These included gas stove use, cigarette smoking, incense and candle burning, practices observed in 82% of monitored sites. AI algorithms, trained on high-resolution sensor data and labelled occupant activity logs, consistently detected sharp, temporally precise spikes in $PM_{2.5}$ and CO concentrations during these events. Most combustion activities occurred during cooking hours (early morning, lunchtime, and evening) or during culturally timed practices such as early morning incense rituals.

In environments with low natural ventilation (defined as air change rates, ACH, below 0.4 h^{-1}), the duration and intensity of these indoor air pollutant spikes were substantial. CONTAM exposure pathway simulations incorporating time-location data revealed that these spikes contributed disproportionately to cumulative exposure: combustion-related emissions were responsible for a median of 48% of long-term PM2.5 dose and 39% of cumulative VOC exposure in such environments. In some cases, $PM_{2.5}$ concentrations reached instantaneous peaks above 250 µg/m³, persisting for 20 to 45 minutes post-combustion in enclosed kitchens or living spaces.

The second major category of indoor air pollutant source identified across the study sites was material-based off-gassing. PMF and AI-based classification both consistently highlighted elevated baseline concentrations of VOCs, particularly formaldehyde and benzene, in buildings containing pressed wood furniture, PVC flooring, polyurethane insulation, or recently painted surfaces. These emissions were not associated with occupant activity and were instead attributed to persistent, low-level volatilisation from composite building materials and furnishings.

Material-based emissions were dominant in 76% of the sites, especially in recently renovated or high-occupant-turnover environments. In these buildings, formaldehyde concentrations frequently exceeded 80 µg/m³, with little temporal variability, suggesting sustained emissions over weeks or months. CONTAM modelling confirmed that material-based sources accounted for approximately 41% of the long-term cumulative VOC dose. The persistence of these emissions, despite intermittent ventilation, suggests that source control (e.g., material substitution or surface sealing) would be essential for long-term risk reduction.

Short-duration behavioural emissions were also captured with high resolution due to the five-minute interval sensor data. These activities included the use of spray-based air fresheners, multipurpose chemical cleaners, and personal grooming products, and were reported in 63% of monitored sites. AI source identification linked these events to sudden spikes in TVOCs and, occasionally, ultrafine PM from propellant mechanisms.

Although the magnitude of these spikes was often substantial, exceeding 500 μg/m³ for $PM_{2.5}$ in some cases, their contribution to annual cumulative dose was relatively low, estimated at approximately 11%. This reflects their short duration and the fact that they were typically localised to bathrooms or small rooms with rapid subsequent dilution. However, the transient nature of these spikes makes them a potentially acute hazard, particularly for children, the elderly, or individuals with pre-existing neurological or respiratory conditions.

While indoor sources overwhelmingly dominated exposure under typical conditions, outdoor pollutant infiltration played a secondary role. Its contribution was more notable in peri-urban sites located near major roadways or construction zones. In these environments, infiltration of traffic-related NO_2 and $PM_{2.5}$ was detectable, particularly under high outdoor concentration episodes. Nonetheless, under typical closed-window conditions, outdoor-derived pollutants contributed only 4–7% of the total indoor air pollutant burden. This finding underscores that mitigation efforts focused on indoor source control are likely to be more impactful than interventions targeting outdoor-air filtration in most scenarios examined in this study.

One of the key strengths of the study was its spatially resolved modelling, enabled by detailed CONTAM simulations and zone-specific sensor deployment. The spatial analysis uncovered significant microenvironmental disparities in pollutant distribution.

Kitchens emerged as consistent hotspots of exposure, accounting for over 60% of cumulative $PM_{2.5}$ exposure in 22 of the 40 sites, even after adjusting for the proportion of time spent in the kitchen. This was primarily driven by lack of mechanical ventilation, poor airflow pathways, and occupant practices such as cooking without opening windows or using exhaust fans.

In 12 sites, enclosed bedrooms with minimal or no mechanical ventilation exhibited formaldehyde levels above the WHO guideline value of 100 μg/m³ for more than 25% of the monitoring period. This is particularly concerning given that these exposures occurred during night time hours when occupants were typically asleep, reducing their physiological defences and potentially increasing neurological susceptibility.

These findings highlight the importance of considering not only average pollutant concentrations, but also when and where high exposures occur, especially in rooms used for prolonged occupancy.

The convergence between AI classification and PMF-based statistical decomposition was high, with agreement on dominant source categories in over 90% of cases. This dual-method confirmation substantially strengthens the validity of the findings.

Where discrepancies existed, such as in differentiating VOCs from synthetic flooring versus cleaning sprays, follow-up manual inspection of time-activity logs resolved classification uncertainty. This synergy between data-driven learning models and unsupervised statistical inference reinforced the study's capacity to draw robust conclusions on indoor air pollutant origin and persistence.

Collectively, these findings confirm Hypothesis 3.1: that the dominant sources of long-term indoor exposure to neurotoxic pollutants in both residential and occupational settings are combustion-related emissions and off-gassing from synthetic materials, particularly in environments with insufficient ventilation.

The combination of high-resolution sensor data, mechanistic simulation via CONTAM, and convergent AI and PMF classification enabled a detailed reconstruction of how and when pollutants accumulate. This nuanced understanding of source-specific and spatially distinct exposure patterns provides a powerful basis for targeted intervention strategies, particularly in protecting biologically vulnerable populations from elevated Parkinson's disease risk.

Effectiveness of Risk-Based Mitigation Interventions

The deployment of tailored IAQ interventions across 20 high-risk sites led to statistically robust and clinically meaningful reductions in cumulative pollutant exposure, providing compelling evidence in support of Hypothesis 3.2. This hypothesis proposed that mitigation strategies tailored to both the dominant indoor air pollutant sources and the biological vulnerability of occupants would yield greater reductions in exposure than generic IAQ advice. The analysis was based on a composite cumulative exposure dose metric that integrated time-weighted concentrations of $PM_{2.5}$ and volatile organic compounds (VOCs), including formaldehyde, benzene, and toluene, measured over a four-week period both before and after intervention.

Across the intervention group, the average reduction in cumulative exposure dose was 43%, a substantial decline considering the high baseline exposure levels that characterised these high-risk sites. These reductions were derived from five-minute resolution IAQ data analysed using generalised estimating equations (GEE), which appropriately accounted for within-household clustering and temporal autocorrelation. The GEE model showed a highly significant interaction effect between time (pre- vs. post-intervention) and group (intervention vs. control), with an interaction coefficient (β_3) of –0.44 ($p < 0.001$).

This finding confirmed that the observed reductions were directly attributable to the tailored intervention strategies and were significantly greater than changes observed in the control group, which received only standardised IAQ educational leaflets. The control group showed an average exposure reduction of just 12%, a figure that was not statistically distinguishable from baseline after adjusting for confounders such as seasonal variability, occupant behaviour, and external environmental conditions.

Intervention effectiveness varied significantly by strategy type, with mechanical ventilation interventions emerging as the most impactful. Sites that received decentralised ventilation units equipped with HEPA and activated carbon filters experienced a 50% average reduction in $PM_{2.5}$ and a 47% reduction in formaldehyde concentrations. These improvements were driven by enhanced removal efficiency and a dramatic increase in effective air change rates (ACH).

Tracer gas validation confirmed that ACH rose from a pre-intervention average of 0.35 h^{-1} to 1.15 h^{-1} post-installation, greatly improving dilution and removal of both particulate and gaseous pollutants. Importantly, the mechanical ventilation units selected for installation were matched to room size, indoor air pollutant profile, and occupant preferences to ensure optimal performance and acceptability.

Source control interventions also delivered strong performance. These interventions targeted high-emission indoor materials, including gas stoves, which were replaced with electric cooktops, and composite flooring materials that were either sealed or covered with low-emission surface treatments. The outcome was a 36% reduction in VOC-related exposure and a 29% reduction in $PM_{2.5}$.

Unlike mechanical ventilation, the impact of source control was characterised by a more consistent and stable reduction in indoor air pollutant concentrations, particularly VOCs, which remained lower throughout the entire four-week post-intervention period. This stability reflects the inherently passive nature of source reduction, once the source is removed or sealed, emissions decline irrespective of occupant activity levels.

Behavioural interventions, though less capital-intensive, provided valuable supplementary benefits. These strategies included occupant education on low-emission product usage, optimal ventilation timing (e.g., cross-ventilating during and immediately after air pollutant-generating activities) and minimising the use of fragrance-emitting products and open-flame incense or candles. Behavioural changes led to an average 21% reduction in cumulative exposure dose.

However, the results in this category exhibited higher inter-site variability, which was attributed to inconsistent occupant adherence and the persistence of deeply embedded habits. Some households reported high compliance and saw reductions nearing 30%, while others exhibited little to no change, indicating the critical role of engagement, feedback, and cultural sensitivity in behavioural programme design.

A key strength of this phase of the study was the stratification of intervention outcomes by biological vulnerability. Vulnerability was previously defined based on a composite index derived from biomarker data indicating systemic inflammation (e.g., CRP, IL-6), mitochondrial dysfunction (e.g., membrane potential, mtDNA copy number), and PD-related genetic variants.

Stratified GEE analysis revealed that biologically vulnerable households saw a 54% average reduction in cumulative exposure dose, compared to a 34% reduction in non-vulnerable households—a statistically significant difference ($p = 0.003$). This result supports the core premise of risk-weighted IAQ management: biologically susceptible individuals benefit more from exposure reductions and are thus the appropriate primary targets of intervention efforts.

These findings have far-reaching implications. First, they validate Hypothesis 3.2 by demonstrating that risk-based, tailored interventions grounded in detailed source attribution and individual vulnerability yield superior exposure mitigation outcomes. Second, the results reinforce the need for integrated IAQ management frameworks that combine engineering controls (e.g., ventilation, source removal) with personalised behavioural guidance.

Third, the study provides a strong empirical basis for advancing precision environmental health approaches, particularly those that combine high-resolution exposure data with biomarker-informed stratification. In practice, this means that IAQ interventions should not be designed solely around architectural typologies or average indoor air pollutant concentrations but should also consider who is being exposed and how sensitive they are to harm.

From a policy perspective, the results highlight the inefficiency of generic IAQ advisories, particularly in high-risk or vulnerable communities. They support the prioritisation of resource-intensive interventions, such as mechanical ventilation and source control, in homes with elderly occupants, individuals with neuroinflammatory conditions, or children with genetic susceptibilities.

In this way, public health initiatives can maximise impact while minimising cost by directing efforts toward those with the greatest capacity for benefit. Furthermore, the use of AI-enhanced exposure models and empirical biomarker data establishes a rigorous foundation for future implementation of smart IAQ interventions, risk stratification tools, and real-time exposure dashboards in residential and workplace settings.

Model-Driven Prediction and Validation

To enhance the practical utility of the intervention findings and extend their generalisability beyond the 40 original study sites, machine learning (ML) models were developed to forecast changes in indoor air pollutant exposure under a range of intervention, architectural, and behavioural conditions. These models were trained on the complete dataset of pre- and post-intervention IAQ measurements, which included high-resolution time-series concentrations of $PM_{2.5}$ and total volatile organic compounds (TVOCs), alongside detailed metadata on building characteristics, ventilation configurations, occupant density, and implemented mitigation strategies.

The overarching objective was twofold: first, to determine whether advanced predictive analytics could accurately forecast exposure outcomes given a known set of contextual parameters; and second, to identify the most impactful intervention combinations for indoor air pollutant mitigation under diverse indoor environmental scenarios. In this way, the modelling effort was intended not only to validate empirical findings but also to build a decision-support tool that could inform scalable, data-driven IAQ management in settings not directly observed during the original field trial.

Among the algorithms tested, random forest regression emerged as the top-performing method. It consistently outperformed linear regression, support vector regression (SVR), and gradient boosting machines (GBM) in both predictive accuracy and model robustness.

The final model was trained on 70% of the dataset and validated on the remaining 30%, using repeated 10-fold cross-validation to avoid overfitting and to ensure generalisability. For $PM_{2.5}$, the model achieved a mean root mean square error (RMSE) of 2.9 μg/m³, while for VOCs, the RMSE was 3.2 μg/m³. These error margins were well within the expected measurement uncertainty range for indoor air pollutant monitoring in real-world settings.

The model's coefficient of determination (R^2), which reflects the proportion of variance explained by the model, reached 0.87 for $PM_{2.5}$ and 0.84 for VOCs, indicating a high degree of predictive fidelity. Importantly, these strong performance metrics held across validation folds and site types, despite the heterogeneity of floorplans, ventilation strategies, occupant activities, and intervention types represented in the dataset. This suggests that the model is well-suited for forecasting IAQ outcomes across a broad range of real-world building scenarios.

One of the most valuable outputs of the random forest model was its ability to produce feature importance rankings. Each predictor variable's contribution to the overall model was assessed using permutation-based importance scores, averaged across 1,000 bootstrap iterations. The most influential variables in predicting exposure reduction were (i) the presence and type of mechanical ventilation with HEPA and activated carbon filters, (ii) the use of source control interventions such as gas stove replacement and flooring encapsulation, and (iii) the baseline (pre-intervention) indoor air pollutant concentration. These findings align with the empirical results from the intervention phase, reinforcing their validity.

The model also revealed important interaction effects. Recursive partitioning within the forest structure showed that the combination of mechanical ventilation and source control consistently outperformed either strategy alone. For example, sites that implemented both approaches achieved predicted exposure reductions nearly 1.7 times greater than those that relied on ventilation alone.

This synergistic effect was observed across both indoor air pollutant classes, $PM_{2.5}$ and VOCs, and supports the design principle that multi-modal mitigation strategies are more effective in complex indoor environments with multiple emission sources.

To strengthen confidence in the mechanistic validity of the underlying simulations, field-based tracer gas validation was performed in a subset of eight intervention homes. A non-toxic, chemically inert tracer gas, sulphur hexafluoride (SF_6), was released in each unit under controlled environmental conditions.

Results showed that measured ACH values deviated from the modelled values by less than 12% across all tested sites. This level of agreement is considered excellent in the context of real-world ventilation measurement and provides strong validation for the ventilation assumptions used in the exposure pathway simulations. In homes where decentralised mechanical ventilation was installed, tracer gas measurements confirmed substantial increases in ACH, from baseline values of 0.3–0.4 h^{-1} to post-intervention levels between 0.9 and 1.1 h^{-1}.

These improvements were not only statistically significant ($p < 0.001$) but also correlated closely with observed reductions in cumulative exposure dose, as confirmed through both direct IAQ sensor measurements and the AI-based model predictions.

Together, these findings confirm the practical value of machine learning models for forecasting IAQ mitigation outcomes. Random forest algorithms demonstrated the capacity to integrate multidimensional environmental, behavioural, and architectural data and produce highly accurate exposure predictions.

This predictive power has broad implications. It enables the development of interactive IAQ risk simulators, early warning systems, and personalised mitigation planning tools that can be deployed in both residential and occupational settings.

Moreover, the demonstrated ability to generalise across diverse sites and intervention conditions strengthens the case for integrating AI-enhanced modelling into municipal health frameworks and building management platforms. By combining empirical monitoring, validated simulation tools, and machine learning-driven forecasting, this approach represents a scalable, adaptive, and evidence-based strategy for managing indoor air pollution and reducing the environmental risk burden for neurodegenerative diseases such as Parkinson's.

6……………………………….

After earning her PhD with distinction, Zahra stood at a turning point. Her doctoral research had already redefined IAQ assessment practices, no longer through static averages and broad compliance indicators, but through dynamic, biologically stratified models that captured how real exposure occurred across time, space, and individual biological vulnerability. She had shown how indoor environments, once dismissed as neutral, could accumulate conditions that increased neurodegenerative disease risk, particularly Parkinson's. Her findings did not just call for improved modelling, they demanded a transformation of practice.

She accepted a postdoctoral fellowship at the National Built Environment Research Centre, a leading institution under the Ministry of Sustainable Built Environment. The Centre had recently received expanded government funding as part of a national strategy to align environmental sustainability with public health resilience.

Zahra's work, grounded in AI-enhanced exposure modelling and health-risk stratification, aligned precisely with the Ministry's forward-looking mandate. With institutional backing and cross-agency cooperation, Zahra now had the platform and resources to move from proof-of-concept to policy-integrated systems.

In the early years of her appointment, Zahra led a high-impact collaboration with the national public housing authority. Her goal was to shift residential IAQ standards from static threshold compliance to predictive, risk-based protocols that integrated biological vulnerability. She adapted her PhD models into operational frameworks that considered emissions, ventilation patterns, and occupancy dynamics while allowing for stratification by age, health conditions, and exposure duration.

These were piloted across multiple housing zones. The results, sharper risk detection, earlier intervention, and reduced cumulative exposure in vulnerable populations, convinced the Ministry to mandate her framework as part of IAQ standards across all new public housing developments.

Her work expanded into the healthcare sector. Zahra launched a government-backed initiative to reform IAQ management across eldercare facilities. Her AI-driven system enabled real-time tracking of indoor air pollutant accumulation in living and treatment spaces, correlating these fluctuations with known sensitivity profiles of residents.

Facility staff received training to interpret the system's adaptive risk forecasts and modify care routines accordingly. The results were tangible: in facilities using Zahra's system, rates of reported respiratory distress and inflammation-related hospital visits declined. For clinicians working with patients at risk of neurodegeneration, Zahra's models began to offer a new lens for understanding environmental triggers of symptom progression.

In her third year at the Centre, Zahra was promoted to staff scientist in recognition of her growing national influence. No longer a fellow, she took on broader leadership responsibilities across inter-ministerial taskforces and international partnerships.

One of her most notable contributions was the development of a national digital exposure mapping platform, which allowed occupants, clinicians, and policymakers to access personalised exposure data linked to residential building types, regional air pollutant baselines, and vulnerability-adjusted thresholds. The platform, launched by the Ministry as a public service tool, became a cornerstone of the country's healthy housing strategy.

Zahra also transformed professional development in the sector. She co-designed and implemented an interdisciplinary certification programme for engineers, public health officials, architects, and community health workers, focused on biologically responsive environmental design and management. Trainees used her AI platform to simulate different building scenarios and evaluate how changes in ventilation, layout, or materials could reduce long-term health risks. Her training modules later became mandatory for accreditation in the national building health assessment scheme.

Internationally, Zahra helped lead the drafting of the country's contribution to emerging global standards on AI-supported IAQ assessment. She insisted on frameworks that were not only computationally robust but ethically grounded, transparent, and adaptable to resource-constrained settings. Under her guidance, her home country became a model for how governments could adopt AI in a way that centred human vulnerability, not just efficiency or compliance.

In the final phase of her five-year tenure at the Centre, Zahra piloted the integration of IAQ risk analytics into primary healthcare systems. Working with neurologists and geriatric specialists, she co-developed digital modules for clinical records that incorporated exposure history, building risk profiles, and early-warning indicators derived from real-time IAQ data. Physicians could now contextualise symptoms, such as fatigue, cognitive fog, or motor slowing, not only in terms of ageing or disease progression, but in relation to long-term environmental stressors.

Throughout these five years, Zahra remained deeply committed to community engagement. She oversaw the translation of her work into publicly accessible formats, such as illustrated guides for safer household practices, school-based indoor air quality learning kits, and interactive platforms that allowed residents to estimate their personal exposure risk and receive tailored recommendations. These were not side projects. They were central to her philosophy that transformation must begin where people live, breathe, and make daily choices.

By the time Zahra completed her fifth year at the Centre, she had done more than extend her PhD. She had transformed it into a national operating system for air-health integration. Her AI-enhanced risk modelling was no longer confined to research articles or simulation dashboards. It now lived in housing policy, healthcare protocols, training syllabi, certification processes, and urban planning regulations. She had built not only tools, but institutions, norms, and capacity—a foundation others could stand on.

When the Ministry offered her a permanent senior position as Director of Integrated Environmental Health Systems, she accepted. However, she remained clear: her work would continue to sit at the convergence of science, governance, and lived experience. The air may have been invisible. But thanks to Zahra's work, its consequences no longer were. And in their place stood something more enduring systems designed not only to monitor, but to protect.

When Zahra stepped into her new role as Director of Integrated Environmental Health Systems under the Ministry of Sustainable Built Environment, she brought with her not only scientific rigour, but also a philosophy rooted in equity and foresight. She no longer needed to prove that indoor environments could influence health. That ground had already been broken. Her mandate now was to ensure that every building system, every assessment protocol, and every professional standard aligned with that truth and acted on it.

One of her first initiatives was the launch of the National Indoor Exposure Intelligence Programme (NIEIP), a comprehensive framework that operationalised AI to predict health risks across diverse building categories, including schools, healthcare facilities, offices, shopping centres, manufacturing plants, and public transit hubs. The goal was not simply to monitor indoor air quality, but to forecast health-relevant risk conditions by continuously analysing environmental data in conjunction with occupancy patterns and vulnerability profiles.

Under her leadership, NIEIP developed sector-specific AI modules that accounted for the unique operational dynamics of each environment. In schools, the models learnt to detect patterns where cleaning sprays used during recess and insufficient cross-ventilation combined to raise VOC and $PM_{2.5}$ levels just as children returned to class.

In hospitals, the models identified microenvironments where sterilisation procedures caused localised pollutant spikes that posed respiratory risks to immunocompromised patients. In manufacturing, Zahra's system differentiated between routine emissions and anomalous patterns that indicated maintenance failures with potential health consequences.

Her most ambitious extension of the programme was the integration of IAQ risk intelligence into national rail systems. Trains, she observed, were mobile indoor environments with high human density, variable ventilation, and transient occupancy, all conditions that traditional IAQ models failed to capture accurately.

Zahra’s team set up smart systems inside the trains that constantly monitored indoor air quality and used AI that could learn and adapt to changing conditions inside the moving carriages. This helped them identify when and where exposure risks were rising and allowed for timely interventions like increasing ventilation or rescheduling cleaning.

These models responded in real time to occupancy surges, weather-induced ventilation changes, and identified specific conditions, such as simultaneous use of air fresheners and closed windows, that sharply increased short-term exposure to irritants. The system provided predictive alerts to train operators, enabling them to adjust ventilation or cleaning schedules pre-emptively.

Beyond deployment, Zahra led the transformation of the assessment methodology itself. Traditional IAQ evaluations had long relied on short-term measurements compared against fixed thresholds, producing risk scores that ignored exposure variability and individual differences. Zahra replaced this with a risk-in-context model, powered by AI, which assessed not only indoor pollutant levels but also time-of-day activity profiles, typical behaviour patterns, and known occupant susceptibilities.

For example, in elderly housing, instead of flagging a unit as compliant or non-compliant, the system produced vulnerability-weighted exposure indices, showing how even moderate indoor air pollutant concentrations during sleep hours could represent a higher cumulative risk for residents with pre-existing neurological or cardiopulmonary conditions. These indices were used to prioritise building interventions where they mattered most, not where the average value looked worst.

Under Zahra’s guidance, this system became the national benchmark. Building owners and facility managers across sectors adopted the framework not as an obligation, but as a risk management tool. Insurance firms began offering reduced premiums for buildings with adaptive exposure systems in place. Employers integrated IAQ-based health risk forecasting into occupational health protocols. Even architects and developers began to consult the NIEIP database to design buildings that minimised indoor air pollutant accumulation under real-use conditions.

Importantly, Zahra ensured that the tools remained transparent and accountable. She established the AI Interpretability and Ethics Board for Environmental Systems, mandating that all algorithmic decisions that shaped exposure alerts or influenced building operation had to be explainable to non-technical stakeholders. Her philosophy never wavered: intelligence was not enough, people had to understand why a risk was flagged, and what action would reduce it.

By the end of her second year as Director, Zahra had done more than integrate AI into IAQ management. She had transformed how risk itself was defined. Not as a number on a dashboard. But as a human-centred condition, dynamic, unequal, and modifiable. And in doing so, she had laid the foundation for a future where indoor environments no longer passively shape health but actively protect it.

When Zahra accepted the full professorship at the country's top university, the National University of Belinburg, the very same institution she had once declined as a young student, she did so not to reflect on a completed legacy, but to advance the very systems she had already transformed.

After seventeen years of public service, including twelve as Director of Integrated Environmental Health Systems and five as a postdoctoral researcher and staff scientist at the Ministry of Sustainable Built Environment, Zahra had fundamentally reshaped how indoor air quality was assessed and managed across the country. Her work had operationalised AI to predict health risks from indoor air pollutant exposure, stratified by biological vulnerability, and made risk intelligence actionable through value-oriented mitigation.

But as she transitioned to the university as Professor of Applied Environmental Health Systems, her mission shifted from scaling implementation to deepening understanding, evolving capability, and seeding the next generation of transformative practice. She did not return to replicate what she had done at the Ministry. She returned to advance it, conceptually, pedagogically, and institutionally.

At the Ministry, she had focused on governance-level system reform: deploying AI tools across buildings and transit networks, shaping national assessment standards, training professionals for operational uptake. These systems forecasted risk, informed policy, and changed behaviour. Now, at the university, her work centred on building the foundational structures, intellectual, ethical, and methodological, that would shape future innovators, educators, and problem-solvers for decades to come.

She established the Applied Intelligence for Indoor Health Lab (AIIHL), which served as the country's first academic centre dedicated to advancing the science and ethics of AI in environmental health. The lab's focus was not merely on model performance, but on how knowledge about exposure and vulnerability could be translated into decisions that maximised health impact per unit of invested effort, time, and cost, the essence of value-oriented problem-solving. While her Ministry work had proven that such AI frameworks could be implemented, the university provided a space to question, iterate, and evolve those frameworks for broader societal application.

The research agenda she led as a professor also expanded the analytical boundaries of her earlier work. At the Ministry, risk was primarily assessed at the building level. In academia, Zahra pushed for the integration of multi-scalar, real-time exposure forecasting—linking household-level pollutant fluctuations with district-wide urban planning variables, energy use patterns, climate-driven ventilation shifts, and socio-behavioural data.

Her team explored how AI could learn not only from environmental inputs, but also from human routines and adaptive behaviour across seasons, cultures, and economic constraints. In short, she transformed static IAQ modelling into a living science of dynamic exposure ecology, rooted in human diversity and systems complexity.

Her contribution to applied learning was equally ground-breaking. While she had developed professional training at the Ministry, at the university she constructed a vertically integrated educational model. Undergraduates, graduate students, professionals, and policymakers all engaged with variations of her problem-based learning curriculum.

Central to this model was the requirement that students use real-world IAQ datasets to co-develop interventions that were not only technically sound, but socially and biologically appropriate, cost-conscious, and explainable to non-expert stakeholders.

In the classroom, Zahra posed design challenges that mirrored unresolved real-world problems: How should a low-cost preschool in a tropical urban slum reduce neurotoxic exposure for malnourished children using no imported technology? How could an office tower's IAQ system be optimised for workers with autoimmune conditions, without raising energy consumption? How can mitigation strategies be prioritised when funding is limited, and vulnerability is unequally distributed? Each assignment fused AI, health science, building physics, ethics, and economics, forcing students to reason beyond their discipline, to think in systems, and to ask: "Whose life improves, and how sustainably?"

Her work in continuing education and training (CET) also reached new heights. At the Ministry, she had designed practitioner courses. At the university, she built a nationally recognised CET academy offering stackable microcredentials in AI-enhanced IAQ forecasting, exposure-risk communication, and value-oriented decision-making.

She embedded simulation environments based on anonymised Ministry case files, allowing professionals to test decisions against predictive models and downstream health impact scenarios. The academy became the primary upskilling pathway for environmental health officers, building engineers, facility managers, and urban planners across the country, and increasingly, the region.

During one of Zahra's advanced CET sessions, a participant raised a question that reflected a common uncertainty in the field, one Zahra had encountered many times before. She welcomed it not as a challenge, but as an opportunity to reaffirm why her work placed equal emphasis on scientific precision and the reality of lived experience.

"[The participant asked]: Professor Zahra, can indoor air pollutants actually lead to Parkinson's disease? I find this connection quite uncertain and somewhat far-fetched. Additionally, since everyone is exposed to some level of indoor air pollution almost all the time, why would only certain people end up developing the disease? Are we sure the link is real? I mean, people get Parkinson's for all sorts of reasons, right? Couldn't it just be a coincidence that some of them were also exposed to pollution?

[Prof. Zahra replied]: These are good questions. Questions that highlight the need to understand what contributes to health risk, and that risk is not only about hazard. We are all exposed to indoor air pollutants all the time, yet not all of us are equally protected from their effects. In this case, the difference lies in health conditions that compromise the brain's defences. Conditions such as diabetes, chronic inflammation, or mitochondrial dysfunction weaken the brain’s defences, especially in the substantia nigra, a region primarily affected in Parkinson’s disease.

The substantia nigra, located in the midbrain, controls movement by producing dopamine and is the primary brain region affected in Parkinson’s disease. People with chronic inflammation (from obesity or autoimmune diseases), mitochondrial dysfunction, type 2 diabetes, or nutrient deficiencies often already have impaired neuron function in this region. These conditions increase oxidative stress, reduce cellular energy, and weaken the brain’s ability to defend and repair itself.

Gaseous pollutants and chemical and biological-based particles further accelerate oxidative stress, especially in individuals with compromised neuronal defences. Thus, long term exposures intensify damage to an already vulnerable substantia nigra with their harmful energy —the hazard. This leads to lower dopamine levels and classic Parkinson’s symptoms: tremors, stiffness, slowness, and balance problems.

[The participant asked]: Professor, among all the possible sources of indoor air pollutants, which ones tend to consistently contribute most to long-term exposure and serious chronic health risks, and why? What approach can be used to effectively predict the risk of developing Parkinson’s disease from long-term indoor air pollutant exposure, particularly when accounting for individual biological vulnerability?

[Prof. Zahra replied]: The main sources of long-term indoor air pollution include combustion activities, like cooking, smoking, and traffic, and emissions from materials such as furniture, flooring, and cleaning products. These are especially problematic because they release harmful pollutants frequently or continuously, increasing chronic exposure risks.

We developed AI tools that outperform traditional models (Time-weighted average model and Multivariate mass balance model) by capturing the dynamic interplay of building-specific, behavioural, and environmental data to accurately predict chronic exposure, biological vulnerability, and PD risk, providing robust insight on exposure-vulnerability interaction, and targeted, risk-based intervention.”

Perhaps most significantly, Zahra used her position to influence educational policy. She led the drafting of a new competency framework for engineering and environmental science education, which was adopted by the university council. It formally required all IAQ-related programmes to incorporate interdisciplinary instruction in biological risk stratification, exposure ethics, and AI interpretability. She argued that modelling without accountability, or intervention without value evaluation, was no longer acceptable in a world where exposure-driven diseases could be silently shaped by daily environmental decisions.

By the time Zahra entered her sixth year as full professor, her influence was visible across every layer of the university’s applied research and education ecosystem. Students cited her coursework as transformative. Faculty members collaborated across departments in ways that were previously unthinkable. Research centres began to explore applications of her value-oriented framework beyond IAQ into water quality, waste systems, and even urban food resilience. Her lab hosted visiting scholars from cities around the world seeking to replicate her methodology.

In leaving the Ministry, Zahra had not stepped away from practice. She had elevated it. She was no longer refining systems on behalf of others. She was equipping thousands to do it for themselves, critically, reflectively, and always with one question at the core: Is this the best use of our effort, for the people it is meant to protect? Through applied research, applied learning, and CET, Zahra had advanced not just a field, but a way of thinking that could never again be unlearned.

Zahra's transformation was not the story of a sudden awakening, nor the triumph of brilliance over challenge. It was the long, deliberate unlearning of a flaw that had been mistaken for virtue. The belief that the more effort, time, money, and complexity poured into a solution, the more valuable it must be. For years, she had clung to that misbelief, mistaking labour for clarity and scale for significance.

However, over time, through trial, failure, listening, and disciplined observation, she reshaped herself. Her value no longer came from what she could show, but from what she could simplify. No longer from what she could build, but from what she could clarify, in context, with care. She had not merely abandoned her flaw. She had rebuilt the mental and emotional infrastructure that once kept it alive.

Her parents had watched Zahra's evolution with quiet pride. Her father, once troubled by her decision to reject the prestigious engineering scholarship, had long since become her most steadfast supporter. As he once admitted, "I used to think you were taking the longer road. Now I see that you simply refused to run in the wrong direction. Her mother, ever observant, often told Zahra's daughters, "Your mother builds things you cannot see until you feel the difference."

Zahra's family life had matured alongside her intellectual path. Her marriage to Akin had become a partnership in both mind and spirit. He understood her commitment not only because he was a clinical epidemiologist, but because he shared her way of thinking: that purpose, not perfection, was the highest goal. Their conversations were often filled with questions about ageing, resilience, probability, airflows, ethics, and parenting.

Their daughters, Temi and Kemi, inherited that spirit. They grew up surrounded by maps of buildings, sketches of human routines, models of airflow, and stories about decisions that protected someone quietly, just in time. When Kemi once remarked, "It's not pretty, but it works, and that's what they needed," Zahra knew something profound had shifted across generations.

Her professional legacy was no longer confined to her titles or innovations. It lived in every building that breathed cleaner air because of her system. It lived in every student who chose to ask why before deciding how. It lived in every public servant who used her training materials to make faster, smarter decisions for vulnerable communities.

Zahra visited her childhood home, where her parents still lived. The house had changed little, but the old forge in the backyard had long been retired, dusty, unused, but intact. Her grandfather had passed some years ago, yet his tools still hung neatly on the wall, as though waiting for his return. The bellows sat folded in the corner, quiet now, but not forgotten.

She stepped inside, the air still holding the faint scent of ash and iron. There, in the same corner where she had once gripped the bellows as a determined six-year-old, she stood for a long moment. The silence was not empty. It was full of memory. Her grandfather's voice returned to her, as clear as it had been all those years ago, "Sometimes, a smaller flame does the work better."

At the time, she had not understood it. She had been too busy proving her strength through how hard she worked, too eager to be seen, to be praised, to be applauded for the scale of her efforts. However, now, standing in that quiet workshop, Zahra finally understood. Her life, her entire career, had become the proof of that single idea. A smaller flame, when focused with purpose, when shaped by care and informed by listening, could soften the hardest metal. It could guide decisions that lasted longer than spectacle ever could. It could protect the people who had always been overlooked by models that averaged them away.

She smiled softly, not because she had conquered the flaw, but because she had turned it inside out. Where once she had wasted effort to feel worthy, she now taught others how to do more with less, not through shortcuts, but through insight. Where once she had equated value with materials and sweat, she now defined it as effectiveness per unit of investment, measured in protection, clarity, and fairness.

Back at the university, Zahra continued to serve, not as someone above the system, but as someone woven into it. She worked closely with ministries, health boards, and construction councils, helping them adopt her models into guidelines that balanced feasibility with urgency. More importantly, she nurtured the next generation of thinkers, people who would not need to unlearn what she once had to.

When she was named the recipient of the International Medal for Transformative Public Health Infrastructure, she stood on stage and accepted it not as an honour, but as an affirmation. "We have spent too long celebrating complexity," she said. "What the world needs now is clarity, delivered with discipline, empathy, and as few wasted steps as possible."

She did not linger in the spotlight. She never had. Later that night, after the ceremony, her daughters asked her what the medal meant. "It means," she said gently, "that someone noticed the fire was smaller, and it still worked." Her husband wrapped an arm around her shoulders. Her mother, seated nearby, nodded in quiet agreement. Her father, standing just behind them, looked at her in silence, his expression full, not of surprise, but of quiet fulfilment. He did not speak, but Zahra knew what it meant. He had seen the flame, all those years ago, and now he saw what it had become.

Zahra no longer needed applause to feel she had done something worthwhile. She needed only to know that the solution had fit the need. That the flame had been just enough. That the air people breathed because of her work was not only cleaner, but safer, fairer, and better understood.

As her story came full circle, it no longer resembled the flaw that began it. It resembled something else entirely a life of deliberate restraint, of critical insight, of value shaped by clarity and purpose. A life that honoured the wisdom of a grandfather who had once tried to teach her something simple. This time, she had listened. **The End!**