# Contents

**Letter from the Editors**  
AMY LIU, EVELYN FAUST, AND KYLE RODRIGUES  

1

**Our Mission Statement**  

2

**Acknowledgements**  

3

**Deconstructing Wahhabism**  
BILAL IRFAN  

7

**Literature Review: The Roots and Clinical Effects of Racial Bias in Medicine**  
KAYLA EGAN  

26

**Extensive Drug-Resistant Typhoid Fever Prevention and Management in Pakistan: A Challenge to Public Health**  
BILAL IRFAN, IHSAAN YASIN, AND DENISE KIRSCHNER  

37

**Evaluation of CDM and RBM Methods to Estimate Small Q-matrices**  
RAPHAEL JEONG-HIN CHIN  

49

**Multimodal Imaging Evaluation of Rabbit Models**  
ASHLEY BROWN, VAN PHUC NGUYEN, IYABODE AJAYI, AND YANNIS M. PAULUS  

59

**An Analysis of SPEEK as a Suitable Alternative for Nafion in a Proton Electrolyte Membrane in PEM Fuel Cells**  
IHSAAN YASIN AND BILAL IRFAN  

68

**Unlocking the Power of Molecular Cloning: Revolutionizing Medical Microbiology Procedures**  
NEELABH DATTA  

86

**Checking the Status: The Evolutionary Explanations and Drug Resistance Prevalence to Dolutegravir for HIV Treatment (A Review)**  
EVAN HALL  

134
A Study on the Effect of Zamzam Water on the Germination and Growth of *Jasminum sambac*  
BILAL IRFAN AND IHSAAN YASIN  
144

Association between Coexisting Hypertension, Dyslipidemia, and Elevated C-Reactive Protein with Cardiovascular Disease in the Health and Retirement Study  
THOMAS KARADIMAS AND HELEN C.S. MEIER  
159

Fitness Costs of Early Life Adversity in Non-Human Primates  
IRENE WANG  
175
Dear Readers,
It is with great pleasure that we present this year’s issue of the University of Michigan Undergraduate Research Journal, which is the result of months of dedication, intellectual curiosity, and hard work from both the talented undergraduate scholars and our staff. Our mission at University of Michigan Undergraduate Research Journal is to give students the opportunity to present and publish their research. In collaboration with faculty members at the University of Michigan, our staff carefully accepts, reviews, and edits submissions to our journal. The collaboration among authors, editors, and faculty members ensures that each piece of research is at its highest quality.

We are excited to showcase research from life sciences, social sciences, math, physics, and humanities in this issue. As editors, it is our privilege to witness the intellectual growth and academic excellence of our contributors. We hope that our journal serves not only as a platform for showcasing individual achievements but also as a catalyst for inspiring others to embark on their own research journeys.

Sincerely,
Amy Liu, Editor-in-Chief (2023–24)
Evelyn Faust, Editor-in-Chief (2023–24)
Kyle Rodrigues, Editor-in-Chief (2023–24)

Contact: Amy Liu <amyzliu@umich.edu>
Evelyn Faust <evfaust@umich.edu>
Kyle Rodrigues <kyrodrig@umich.edu>

https://doi.org/10.3998/umurj.5499
We are a student-run, non-technical research journal. Our mission is to build connections between undergraduates, graduate students, and the public, as well as among the different academic disciplines through the publication of non-technical articles in all fields of research. Our goal is to tap student creativity and effort in all academic disciplines through a peer-edited, faculty-reviewed electronic and print publication. We hope to inspire interest in research in the student body, faculty, and the campus community.

Contact: UMURJ Editors <umurj-editorsinchief@umich.edu>
Our Sponsors

UROP

The Undergraduate Research Opportunity Program (UROP) creates research partnerships between undergraduate students and University of Michigan researchers and community partners. All schools and colleges are active participants in UROP, which provides a wealth of interesting research topics for program participants. UROP started with 14 student/faculty partnerships in 1988 and has expanded to include more than 1300 students and 800 faculty researchers.

The purpose of UROP is to support the mission of the University of Michigan (U-M) by engaging undergraduate students in research, scholarship, and creative inquiry. UROP encourages students toward a lifelong appreciation for discovery and social benefits of diversity, for building understanding across differences, and for critically examining information in the world around them. By providing opportunities for students to work on research and creative projects with U-M researchers and community partners in all academic disciplines, UROP creates conditions for an equitable, inclusive, and supportive educational environment where every person feels valued and has an opportunity to add significance.

UMOR

With 19 schools and colleges and more than 200 centers and institutes, the University of Michigan (U-M) is leading in research in virtually every major area of science, engineering, medicine, business, arts and humanities, public policy, and education.

The U-M Office of Research catalyzes, supports, and safeguards research across all three campuses so that students, faculty, and staff can advance

---

Contact: UMURJ Editors <umurj-editorsinchief@umich.edu>

https://doi.org/10.3998/umurj.5501
knowledge, solve challenging problems, create new scholarly works, and enhance quality of life.

As the nation’s largest public research university, the U-M Office of Research works with leaders across the campus to cultivate interdisciplinary research, diversify sources of funding, improve the efficiency of research operations, and strengthen partnerships with industry, government, and peer institutions worldwide.

UM Library

Consistently ranked as one of the top ten academic research libraries, the University of Michigan Library makes available an extraordinary array of resources and services. Our mission is to support, enhance, and collaborate in the instructional, research, and service activities of the faculty, students, and staff, as well as contribute to the common good by collecting, organizing, preserving, communicating, publishing, and sharing the record of human knowledge. Expert staff are committed to helping learners and scholars tap into the full potential of these informational resources and to providing a full spectrum of assistance for research and teaching. We collaborate with students at every step in their educational career and engage with faculty and graduate students’ research. We provide a hub for scholarly publishing via Michigan Publishing that is responsive to scholars. Connecting students, the public, and multiple disciplines through the publication of non-technical articles is an opportunity to showcase student excellence and public engagement. We are delighted to share our expertise and resources with campus partners and students as part of the University of Michigan Undergraduate Research Journal.

Thanks to our Faculty Reviewers

We thank the following faculty for their contributions to ensuring quality work in the undergraduate research journal and providing guidance through feedback for all of our journal submissions.

Social Science

• Dr. Andrew Brouwer, PhD
• Dr. Christopher Fariss, PhD
Life Sciences

- Dr. Gary Huffnagle, PhD
- Dr. Rosa Munoz, BS
- Dr. James Byrd, PhD
- Dr. Guan Xu, PhD
- Dr. Tomoyuki Marukami, PhD
- Dr. Jason Siegel, PhD
- Dr. Lynn Carpenter, PhD

Humanities

- Dr. Sarah Buss, PhD
- Dr. Juan Cole, PhD

Math and Physics

- Dr. Pariksheet Nanda, PhD

Special Thanks to

- Luciana Aenășoaie
- Amanda Peters
- Doreen Bradley
- Jason Colman
- Charles Watkinson
- Sean Guynes
- Lauren Stachew
- Marisa Mercurio
- Rebecca Welzenbach
- Eric Shaw
- Alex Piazza
- Nick Wigginton
- Jimmy Branco

... without whom this colossal effort would not have been possible. Thanks to each of you for your outpouring of support along the way. We would also like to thank everyone at the Department of Communication Studies, the Physics Department, the Department of Classical Studies, the Philosophy Department, and the English Department for their generous contributions to the journal.
Co-Editors in Chief

Amy Liu
Evelyn Faust
Kyle Rodrigues
Amogh Angadi
Daisey Yu

*Life Sciences*

Kelly Deng, Associate Editor
Jasmine Wisniewski
Namratha Nelapudi
Minseo Lee
Shriya Jaggi

*Physics/Math*

Rafee Mirza, Associate Editor
Azeem Saifee

*Social Sciences*

Grace Abbud, Associate Editor
Serina Jiang
Dahika Ahmed

*Humanities*

Shannon Liang, Associate Editor
Shichi Dhar
Matthew Lee

Contact: UMURJ Editors <umurj-editorsinchief@umich.edu>
Deconstructing Wahhābism

Bilal Irfan

Introduction

When addressing the rise of Wahhābism, particularly its distinction from the Ḥanbalī school of jurisprudence, one must dive into the prevailing narratives that were propagated by the former that enabled it to take to this new identity. In the contemporary era,¹ many Wahhābis self-identify as Salafis,² due to the negative connotations associated with the former term in alleged links to extremism or fundamentalism.³ Some have chosen to label it (the movement) as the Najdi Dawah,⁴ in the view that Muslims should not use the word “Wahhābi”

₁. Lacey, The Kingdom: Arabia and the House of Sa’ud, 56.
₃. MacFarquhar, A Few Saudis Defy a Rigid Islam to Debate Their Own Intolerance.
₄. Commins, The Wahhabi Mission and Saudi Arabia, 41 (“Official Egyptian correspondence expressed sectarian hostility to the Najdi reform movement”); Ibid., 141. (“Nevertheless, significant differences separate the Najdi movement from the modern reviver agenda because the former stemmed from Muhammad ibn Ad al-wahhab’s distinctive views on doctrine, whereas the

Contact: Bilal Irfan <birfan@umich.edu>

https://doi.org/10.3998/umurj.5503
in a derogatory sense, as is prevalent,\textsuperscript{5} due to its similarity with the name of Allah (al-Wahhāb).\textsuperscript{6} Wahhābī is used as an identifying feature, thus, for those Salafis who may call themselves “simply Muslim” (\textit{muslimūn}) or rightly guided monotheists (\textit{muwahhidūn})\textsuperscript{7} and, at times, are self-proclaimed to be within the parameters of the Ḥanbalī \textit{madhhab} (a school of thought) in matters related to \textit{fiqh} (Islamic law), which, according to critics, follow the teachings of Muhammad ibn ‘Abdu’l-Wahhāb.\textsuperscript{8} The 18th-century preacher and theologian articulated a drastically different approach,\textsuperscript{9} constituting a separate school of thought, in regard to methodology, the use and parameters of \textit{qiyās},\textsuperscript{10} \textit{ijtihād},\textsuperscript{11} and \textit{ijmā'},\textsuperscript{12} from the Ḥanbalī school they claim to be members of and the other three major Sunnī schools of fiqh.\textsuperscript{13} Thus, Wahhābism could be classified as being akin to a new \textit{madhhab}, though such a label would find critics among Wahhābis.\textsuperscript{14} Despite their internal resistance, they would fall under the paradigm of a \textit{madhhab}, on account of either their belief of following only the true literalist words of the

\textsuperscript{5} Bilal Philips, \textit{The Evolution of Fiqh (Islamic Law & The Madh-habs)}, 135. (“As a corollary to these beliefs, it has been stated that anyone who dares openly to deny the infallibility of all four Madh-habs or the obligation to follow one to these Madh-habs is considered an accused innovator and apostate. In the 20th century the most commonly used epithet for describing such an apostate has been the label Wahhābī.”)

\textsuperscript{6} Qadhi, \textit{On Salafi Islam}, 3.

\textsuperscript{7} Mattaret.al., \textit{The Encyclopedia of the Modern Middle East and North Africa}. (“Definition of Muwahhidūn: The movement was started by a religious scholar from Najd (Saudi Arabia), Muhammad ibn Abd al-Wahhab (1703–1792), schooled by ulama (Islamic clergy) in what is now Iraq, Iran, and the Hijaz (western Arabia).”); Lacey, \textit{The Kingdom: Arabia and the House of Sa’ud}, 56.

\textsuperscript{8} See, for example, Sulaymān ibn ‘Abdi’l-Wahhāb’s critique of his brother’s teachings, Al-Ṣawā’iq al-ilāhiyyah fi al-radd ‘alā al-Wahhābiyyah, as quoted in Zargar, \textit{Origins of Wahhābism from Hanbali Fiqh}, 66.

\textsuperscript{9} De Bellaigue, \textit{Cairo}, 15–16.

\textsuperscript{10} Deductive analogy.

\textsuperscript{11} Independent legal reasoning.

\textsuperscript{12} Scholarly consensus.

\textsuperscript{13} The Shāfiʿī, Mālikī, and Ḥanafī schools are considered (alongside the Hanbali) to constitute Sunni Islam per the 2005 Amman Message.

\textsuperscript{14} Bederka, \textit{Wahhābism and Boko Haram}. (“Followers of Wahhābism will always agree that they are Sunni Muslims, but emphatically reject the label of ‘Wahhābist’. […] Calling them Wahhābis implies that they learned ideas from a man—Muhammad ibn Abdul Wahhab—instead of the Qur’an and Sunnah, the two great sources of Islam.”); Wiktorowicz, \textit{Anatomy of the Salafi Movement}, 235; Qamar, \textit{Wahhābism, Understanding the Roots and Role Models of Islamic Extremism}. 
salaf (the pious predecessors) or their apparent reliance on taqlisid.15 What sets them apart from the traditional four schools of fiqh, or madhhabs, is the implicit,16 and at times explicit, rejection of the authority and the legitimacy of the other madhhabs.17 Much of the early focus of Ibn Wahhâb’s teachings, ascertained by the numerous books he wrote on the subject,18 had been on condemning the practices that many Muslims conducted during their visitation of the graves of saints, companions, or prophets. His view was radically different from the consensus of preceding Islamic jurists, given he considered practices such as kissing or wiping graves to be actions of a people who are astray from the truth. He went so far as to pass takfîr on those who sought aid of any kind from a deceased saint or prophet,19 considering them to have left the fold and protection of Islam and thereby deserving of death. Crucially, this was a violent chain of thought, in that it utilized certain non-violent actions to justify the classification of another recognized Muslim to be an apostate. This position was in stark contrast to the consensus established by the other four schools of fiqh and the Ḥanbalî school that many Wahhâbis claim to be members of. However, there needs to be a concentrated attempt at analyzing the development of the stances on Ziyârah by Ḥanbalî and other Muslim jurists who paved the way for this blanket takfîr passing.

Naturally, it comes to mind the underlying reasons and methods as to how and why Ibn ‘Abdu’l-Wahhâb was able to amass a following to propagate his view on the essence of tawhîd.20 It would be a mischaracterization to claim that this remote Najdi scholar went about to formulate new categories of Islamic theology (apart from hailing within the Atharic creed) that may have caused further controversy. Rather, the inauguration of debate centered around his person, and ideas stemmed from how rigid and unbending they were in practice and interpretation. They shattered the confines of what it meant to be truly Muslim, with

15. For all the criticism leveled by Wahhâbi scholars against the followers of the four schools of Sunni fiqh for considering their scholars near-infallible, many Wahhâbis have been noted to consider the words of scholars such as Ibn Taymiyya, Ibn ‘Abdu’l-Wahhâb, and contemporary Al-Fawzan (and their interpretations) to be supreme in Islamic law.
16. Abou El Fadl, The Great Theft: Wrestling Islam from the Extremists, 57. (“The Wahhâbis used to label themselves al-Muslimun (the Muslims) or al-Muwahhidun (the monotheists), intimating that those who did not accept their creed were neither Muslims nor monotheists”)
17. Algar, Wahhâbism: A Critical Essay, 1–2. (Wahhâbis themselves prefer the titles al-Muwahhidun or Ahl al-Tauhid, “the asserters of the divine unity”. But precisely this self-awarded title springs from a desire to lay exclusive claim to the principle of tawhid that is a foundation of Islam itself; it implies a dismissal of all other Muslims as tainted by širk. There is no reason to acquiesce in this assumption of a monopoly, and because the movement in question was ultimately the work of one man, Muhammad b. abd-al-Wahhab it is reasonable as well as conventional to speak of “Wahhâbism” and Wahhâbis.)
19. To excommunicate a person from the fold of Islam, to label them a Kafir (disbeliever).
20. Oneness of God.
the collateral effects of this line of questioning being reminiscent of the early
days of Islam.

Something that Ibn ‘Abdu’l-Wahhāb frequently decried was that a large
chunk of Muslims during his time and preceding him had all but left the fold of
Islam, due to their failure in professing true tawhīd given their reliance on saints
and righteous figures to intercede on their behalf. He makes a case akin to that
of Ibn al-Qayyim, who was far more willing to make takfīr than many of his con-
temporaries, on the basis of the veneration of Sunnī mystics and saints, the belief
of intercession, and reliance on objects and praises for protection. Ibn Suḥaym
penned a famous epistle to Muslim jurists, scholarly figures and the public,21
wherein he narrates some of the perceived innovations that the Wahhābi sect has
brought about,22 including the claim that Muslims have ceased to be members
of the Islamic community for over six centuries.23 Yet, Ibn ‘Abdu’l-Wahhāb’s under-
standing of tawhīd ran parallel in many ways to the Taymiyyan understand-
ing where it was separated largely into two distinct categories, one being the belief
of the existence of one and only God and the other consisting of the actions and
unequivocally of God’s nature, power, and might that cannot be manifested or
by human means.

The militant groups in the recent era that lay their claim to power and spiri-
tual and political authority cite scripture and Ibn Taymiyya extensively as
the backdrop toward presenting a justification to other members of the Sunnī
Muslim community. A completely baffling problem presents itself regarding
the role of Ibn Taymiyya in shaping Wahhābist ideology or militantism, given
that Muhammad ibn ‘Abdu’l-Wahhāb frequently draws upon Ibn Taymiyya as
the source and credibility for his arguments, especially when labeled to be one
without formal knowledge or scholarship. His Ḥanbalī opponents have often
asserted that he misunderstood and misrepresented, at times deliberately, Ibn
Taymiyya’s positions on key issues. Interestingly enough, his Shāfiʿī opponents
have no qualms about grouping him with Ibn Taymiyya, forwarding the notion
that such heretical ideas have their roots in an already-defamed and distrusted
theologian in Ashʿarī circles.

Ibn Taymiyya’s famous fatwā,24 which stipulated the legality of declaring
takfīr on the Mongols for alleged apostasy and failing to hold on to the tenets
of Islam, has been used systematically to justify a call to arms against Sunnī Muslim
governments that fail to follow a specific, rigid interpretation of Islamic law and
are viewed as having been corrupted by Western customs. Although much of
these underpinnings became vital to the propagation of Wahhābism, including

21. ʾilā mana yaṣīlu ilayhi min ʿulamāʾ al-Muslimīn.
22. min bida’ihi waḍālalātihi.
23. al-nās min sitt miʾat sana laysū ʿalā shay’
24. A legal opinion made by a muftī or a high-ranking Muslim jurist.
the “defense” of the Saudi state, the political implications of Ibn Taymiyya’s works were not the foundational basis of Muhammad ibn ‘Abdu’l-Wahhāb’s message. In his eyes, the duration since the final revelation gave more room for heresy and deviation from the correct method to worship God, as displayed by the daily occurrence of the visitation of graves and the veneration of saints, which was considered to be impeding upon the meaning of being a Muslim, one who submits. *Iḥtiḍāʾ al-ṣirāṭ al-mustaqim li-mukhālafat aṣḥāb al-jahim* (Requiring the Straight Way against the Adherents of Hellfire), penned by Ibn Taymiyya, serves as the much sought out scholarly justification for Ibn ‘Abdu’l-Wahhāb to advance his *takfīr* on his contemporary Muslims, be they laymen or those in positions of authority, for failing to quell what he considered to be manifestly un-Islamic practices. This paved the way for the ensuing struggle between proponents and detractors of Wahhābism who attempted to frame their understandings of Ibn Taymiyya and Ibn Qayyim al-Jawziyya’s works.

Muhammad ibn ‘Abdu’l-Wahhāb reopened a contentious topic regarding the right to refer to the Qurʾān and hadīth directly, paving the way for even more scholarly feuds and dissenting opinions. It crossed the confines of scholarship, oral and written arguments, into one that engulfed the livelihood of every Muslim in a world fragmented across social and political boundaries in an attempt to ascertain what constituted *tawḥīd* and *shirk*.

**Taymiyyan Origins**

Muhammad ibn ‘Abdu’l-Wahhāb definitely gained a wide breadth of interaction with the works and interpretations of Ibn Taymiyya and Ibn Qayyim al-Jawziyya due to his origins in Najd. The region, which has remained poorly defined, stayed largely out of the clutches of Ottoman sovereignty for the better part of the second millennium. Hanafī jurists maintained formidable control over Ottoman functionaries, given it was the preferred *madhhab* for many governments, in part due to its flexibility and emphasis on public welfare in making judgments. However, in the midst of inner Arabia, Ḥanbalī thought remained as the predominant force in matters of jurisprudence and theology, dated to as early as the 8th/14th century. Ibn Taymiyya and his student Ibn Qayyim al-Jawziyya were considered great figures in scholarly circles in al-‘Aḥsāʾ and Najd, and thus, Ibn ‘Abdu’l-Wahhāb’s perceived mischaracterization of their

---

25. It is abundantly clear that the various incarnations of the Saudi state spearheaded by members of the eponymous dynasty have conducted offensive operations under the guise of defending true Islam.

opinions was shut down by many Ḥanbalī contemporaries of the era. Although geographic proximity definitely did play a role in the refutations made against Ibn ‘Abdu’l-Wahhāb’s ideas, its relevance in terms of whose duty it fell to curtail this thought process was also present. Much of the early refutations to Wahhābism seem to have origins in Shāfi’ī or Ḥanbalī critics, which can be attributed to some of the historical narratives surrounding Ibn ‘Abdu’l-Wahhāb’s messages. Shāfi’īs had faced a plethora of issues in tackling Ibn Taymiyya’s rejection of the authority of Ash’arī theology and viewed him with much distrust in terms of his ability to handle issues pertaining to the classical traditions of Sunnī Islam. His views on the veneration of saints were a serious point of contention for other Sunnī scholars, relegating him a position of noteworthy mention while living in much infamy until the rise of contemporary Wahhābism. Many Ḥanbalīs, however, recognized the pivotal role Ibn Taymiyya played in reshaping their school of jurisprudence and lending it an authority to talk on a wide range of issues and thus were quick to rebut much of Ibn ‘Abdu’l-Wahhāb’s preaching. Due to the minority position that the Ḥanbalī school occupied for much of the better part of Islamic history, many of its jurists began to shift the internal consensus to accept methods of ruling that were accepted by the other schools of fiqh. In that vein, criticism of Ibn Taymiyya’s ideas can be found in abundance in Ḥanbalī texts, yet he was viewed with a sense of awe and esteem. Such was the compelling force for Ibn ‘Afāliq, who practiced as an expert of his school in the territory of al-Aḥsā’, to bandwagon with other Ḥanbalī scholars in guarding Ibn Taymiyya and his student’s legacy, even if it meant essentially repudiating the tone of his earliest refutation of Ibn ‘Abdu’l-Wahhāb, which highlighted Taymiyya’s erred views on Ṣūfī saints.

Al-Qabbānī remained one of the most steadfast opponents of Ibn ‘Abdu’l-Wahhāb, which can easily be attributed to his Shafi’ī allegiance and established negative perception of Ibn Taymiyya. In an open message addressed to Ibn ‘Abdu’l-Wahhāb, he accuses him of being led astray by Ibn Taymiyya and being guided by his reprehensible views on tawassul (a means to obtain something by, in this case favor with God) and istighātha (seeking out the dead for assistance). He claims that Ibn ‘Abdu’l-Wahhāb has made the egregious error of following suit in Ibn Taymiyya’s positions that were among his most volatile and highly critiqued. It is equally clear that al-Qabbānī makes no effort to

---

29. There is frequent mention in *Fāṣl al-khiṭāb* denoting that Ibn ‘Abdu’l-Wahhāb had conducted what constituted reprehensible imitation (al-ṭaqlīd al-radī). See, for example, ibid., f. 52b (four instances); idem, *Kashf al-ḥijāb*, ff. 107b (two instances), 109b (two instances), as quoted in, Bunzel, *Manifest Enmity*, 115; fa-min al-ma lām alladhī lā mirjāta filih annaka qallada ‘bn Taymiyya fimā ṣaddathu asāfīn al-ʿilāmā ‑a-la-lāmin [read: min] ḥafawātihī wa-khurāfātihī wa-tabi ṭahu fi maqālatihī
hide his disdain for the perceived crime Ibn ‘Abdu’l-Wahhāb had committed, by making mention of how Ibn Taymiyya was excommunicated from the fold of Islam by numerous scholars of the highest authority of this time. It was done no doubt to make it clear to the public audience of scholarship and others that Ibn ‘Abdu’l-Wahhāb’s actions of takfīr passing, more or less, made him too a prime victim of it in this circular paradigm. Yet, al-Qabbānī’s criticism for Ibn ‘Abdu’l-Wahhāb was not only reserved to the realm of refutation and attributing Ibn Taymiyya’s heretical ideas to this new preacher but also crossed the boundaries into an attack on his academic and scholarly standing. This is apart from the long-established tradition-like critiques of Ibn ‘Abdu’l-Wahhāb’s peculiar lack of education in matters of complex fiqh and theology. He launches a claim that Ibn ‘Abdu’l-Wahhāb copied verbatim proofs, textual evidence, and phrases from a fatwā of Ibn Taymiyya. In attempting to draw a comparison between the Kalimāt published by Ibn ‘Abdu’l-Wahhāb and the Majmūʿ fatāwā, al-Qabbānī makes it clear to the learned audience of Ibn Taymiyya’s works that he has not done an extremely thorough check into what actually constituted Ibn Taymiyya’s preaching. For if he had, he would have recognized perhaps quite quickly that the parallelisms he drew between Ibn Taymiyya and Ibn ‘Abdu’l-Wahhāb’s epistle were tangential to say the least. Thematic similarities cannot be ignored, but rephrasing and the expanding upon the ideas could not be classified easily as simply the sort of copy-pasting al-Qabbānī was alleging. Of course, given Ibn Taymiyya’s glory was largely confined to the Ḥanbalī school, al-Qabbānī’s knowledge of him was based upon many of the refutations of Ibn Taymiyya or those by other refuters of Ibn ‘Abdu’l-Wahhāb who drew a similar comparison. Al-Qabbānī was clearly driven by a sense of extreme hatred for the likes of Ibn Taymiyya and those who shared his doctrines, claiming that the numerous

---

30. min hafawātihi ‘illati lā yanbaghī dhikruhā . . . qālū bi-kufrihi bi-sababihi (“one of his unmentionable faults . . . because of it they called for his excommunication.”), ibid., ff. 108b-109a; cf. idem, Faṣl al-khiṭāb, f. 52b., as quoted in, Bunzel, Manifest Enmity, 115.

31. It is considered one of his canonical fatwās and can be found even in the famous collection of Ibn Taymiyya’s writings known as Majmūʿ fatāwā.

32. Compare al-Qabbānī, Faṣl al-khiṭāb, ff. 31b, 45a and Ibn Taymiyya, Majmūʿ fatāwā, 27:77, 82, respectively, as quoted in, Bunzel, Manifest Enmity, 116.

33. Within al-Qabbānī’s criticism of Ibn ‘Abdu’l-Wahhāb, in attempts to prove similar doctrines between Ibn Taymiyya and the new preacher, he quotes extensively from the refutations by Shāfīʿī scholars, such as ‘Abd al-Ra’ūf al-Munāwī, Taqī al-Dīn al-Subkī, and Ibn Ḥajar al-Haytamī. Despite Shāfīʿī taking the lead in anti-Taymiyyan rhetoric at the time, a Ḥanafī scholar such as Shihāb al-Dīn al-Khafājī has influenced al-Qabbānī’s refutation.
refutations of Ibn Taymiyya was only a positive thing and something merited on account of the preaching that he and his pupils conducted. He frequently bands together Ibn Taymiyya and his contemporary followers into a group worthy of condemnation, as well as those emulating him to this day. In one such remark, he laments that no one would be worthy to be saved on the Day of Resurrection save for Ibn ‘Abdu’l-Wahhāb and Ibn Taymiyya, as well as those few, 12 in all, who do follow him. Nevertheless, al-Qabbānī cannot be said to have been completely blinded by his dislike for Ibn Taymiyya, given that he acknowledged at times the differences in the approach, whereby Ibn ‘Abdu’l-Wahhāb was labeled as even more extreme in takfīr passing. Without going into length at it, he was able to note that the new movement, in essence, declared tawassul and istighātha as acts of kufr that made one leave the fold of Islam. Yet, for al-Qabbānī, similar to his peer Ibn ‘Abd al-Laṭīf, Ibn Taymiyya was still the source to blame at the bottom. Nevertheless, al-Qabbānī cannot be said to have been completely blinded by his dislike for Ibn Taymiyya, given that he acknowledged at times the differences in the approach, whereby Ibn ‘Abdu’l-Wahhāb was labeled as even more extreme in takfīr passing. Without going into length at it, he was able to note that the new movement, in essence, declared tawassul and istighātha as acts of kufr that made one leave the fold of Islam. Yet, for al-Qabbānī, similar to his peer Ibn ‘Abd al-Laṭīf, Ibn Taymiyya was still the source to blame at the bottom.

---

34. al-Qabbānī, Kashf al-ḥijāb, f. 215b, (“wa-huwa ḥaqīq bi-dhālika (He [i.e., Ibn Taymiyya] deserves it”), as quoted in, Bunzel, Manifest Enmity, 117.

35. al-Qabbānī, Kashf al-ḥijāb, f. 122a, (“Ibn Taymiyya wa-talāmidhatuhu (Ibn Taymiyya and his students)”), as quoted in, Bunzel, Manifest Enmity, 117.

36. al-Qabbānī, Faṣl al-khiṭāb, f. 122, (“Ibn Taymiyya wa-man qalladahu (Ibn Taymiyya and his emulators”)”, as quoted in, Bunzel, Manifest Enmity, 117.

37. See Bunzel’s comment on this at length explaining the number 12, Bunzel, Manifest Enmity, 117. (“Why the number twelve is not clear to me.” The comment is surrounded by quotations from several ḥadīth, describing how the Prophet will lead mankind to Paradise on the Day of Judgment. Al-Qabbānī accuses Ibn ‘Abdu’l-Wahhāb of seeing himself in this prophetic role. For these ḥadīth, see Abū Nu’aym al-Iṣbahānī, Dalā’ il al-nuwaywa, ed. Muhammad Rawwās Qal’ajī and ‘Abd al-Barr ‘Abbās, 2 vols. (Beirut: Dār al-Nafā’ is, 1406/1986), 1:65–66, nos. 23–25. Ibn ‘Abdu’l-Wahhāb seems to refer to this comment in one of his letters, saying of al-Qabbānī, “He writes in his work that he only opposes in his work Ibn Taymiyya, Ibn al-Qayyim, and ten others, I being the tenth of them, and the total being twelve” (wa-yaqūla fi taṣnīfihun innahu lam yakhlīf fi taṣnīfih illā ‘in Taymiyya wa bn al-Qayyim wa-’ashara anā ‘āshiruhum wa’l-jamī` iθnah ‘ashar). See Ibn Ghannām, Tārīkh, 1:425 (letter to Aḥmad ibn Ibrāhīm). The comment is noted in Cook, “Origins of Wahhābism,” 200n87.)

38. zidta ‘ilayhi bi-kawn al-tawassul wa-l-istighātha kufran qāti’ an lil-Islam (“you have surpassed him in making tawassul and istighātha an act of unbelief terminating one’s Islam.”), Kashf al-ḥijāb, f. 108b, as quoted in, Bunzel, Manifest Enmity, 118.

39. al-Qabbānī, Faṣl al-khiṭāb, f. 23b, as quoted in, Bunzel, Manifest Enmity, 118.

40. wa-lā yuqhtarrar bi-man istanada ilayhi ḥādāh ‘l-jāhil al-kārī min al-jahālāt fi muurr [read: amarr] almanāhil mithl Ibn Taymiyya wa-man nahā nahwahu . . . fa-aṭlaqa a’imma a’lām fihi ‘l-alsina . . . ‘ulamā’
‘Abd al-Wahhāb ibn Aḥmad Barakāt al-Ṭandatāwī, another Shāfi’ī scholar, in his Kitāb rad’ al-dalāla wa-qam al-jahāla, goes about mentioning one short of a dozen well-known jurists of the four schools of Sunnī Islam, making mention of Ibn Qayyim al-Jawziyya, Ibn Taymiyya, and Ibn al-Jawzi as prominent figures in the Ḥanbalī school.41 Al-Ṭandatāwī is notably different from many of the early Shāfi’ī refuters of Wahhābīsm, in that he does not go out of his way to condemn Ibn Taymiyya, as some of the other scholars hailing from modern-day Iraq did.

It is interesting to note that many later Shāfi’ī critics of Wahhābīsm seem to rely on juristic and theological viewpoints held by their schools and the Ash’arīs to counter this new form of Aḥarī-based textualism. It is attributable, in part, due to the staunch defenses of Ibn Taymiyya presented by Ḥanbalī scholars who render the likes of Sayyid al-‘Alawi ibn Ahmad ibn Ḥasan ibn ‘Abdullah ibn ‘Alawi al-Ḥaddād and al-Rāwī to not draw as many similarities between Taymiyyan and Wahhābīst philosophy.42 Al-Ḥaddād authored Miswbah al-Anam wa Jala’ az-Zalam fi Radd Shubah al-Bid’i an-Najdi allati Adalla biha al-‘Awamm in 1325/1907 and does not expand upon a corrupt Taymiyyan undertone.43 Similarly, Ahmad ibn Zayni ad-Dālān who served as Grand Mufti of Mecca authored a treatise known as Fitnat al-Wahhabiyya in 1878 and did not mention the Taymiyyan influence in this famous work nor at length in the chapter regarding the fitna of Wahhābism in Khulaswat al-Kalam fi Bayan Umara’ al-Balad al-Haram.44

While Ḥanafī and Mālikī refuters who were contemporaries of Ibn ‘Abdu’l-Wahhāb were scarce, Ḥanbalī jurists from Najd and inner Arabia were relentless. They spoke with a similar vehemence over the alleged misreading and misrepresentations of Ibn Taymiyya’s works, with his own brother Sulaymān ibn ‘Abd al-Wahhāb being a vocal critic of him till his region came under Wahhābī influence and he was prosecuted into silence. He claims that his brother plucked what he wanted from Ibn Taymiyya and Ibn al-Qayyim’s works to apply how he pleased to his own doctrinal advancement. Something worth mentioning is that in al-Ṣawā’iq al-ilāhīyya, he does not explicitly state the names

42. al-Ḥaddād addressed here is the great-grandson of Abd Allah ibn ‘Alawi al-Ḥaddād, the famous Yemeni scholar.
43. See name translation, The Lamp of Creatures and the Illumination of Darkness Concerning the Refutation of the Errors of the Innovator from Najd by Which He Had Misled the Common People.
44. See name translation, The Wahhabi Fitna, The Summation Concerning the Leaders of the Holy Sanctuary.
of the aforementioned scholars, but the context allows for it to be read so that one can infer that. Respect for Ibn Taymiyya can be seen through the reference to Ibn Taymiyya as a Shaykh al-Islām in a number of Ḥanbalī refutations, with Ibn ʿAfāliq considering Ibn al-Qayyim to be among the greatest scholars of the Islamic tradition in his letter to ʿUthmān ibn Muʿammar. Ibn ʿAfāliq further writes in an open message to the aforementioned individual that he will go about to explain the true meaning of the words of Ibn Taymiyya and Ibn al-Qayyim due to how badly distorted they have become under Muhammad ibn ʿAbdu'l-Wahhāb. He accuses Ibn ʿAbdu'l-Wahhāb of simply finding whatever in a certain chapter of Ibn Taymiyya’s works he likes to adopt, while disregarding the rest including any peculiar points or restrictions on the unfiltered mode, in a quest to make takfīr easier upon all. Among the students of Ibn Fayruz was a particular Najdi known as Al-Razīnī who embarks on the practice of name-calling Ibn ʿAbdu'l-Wahhāb with epithets, such as ṭāghūt, in explaining how he has lulled a large chunk of people to misrepresent Ibn Taymiyya and his student’s works.

Ṭāghūt has been used in Islamic discourse to talk about a tyrant or someone who has reached the pinnacle of oppression and cruelty. It has been used to refer to those who worship beings or ideas other than God, leading people astray from the height of their power. It has been said to be a creature who commits the crime of rebelling against God and then defying his will. While its Arabic

45. akhadhtum min qawlihim mā jāza lakum dūn ghayrihi (“You have taken from their words what is agreeable to you to the exclusion of what is not.”), Sulaymān ibn ʿAbdal-Wahhāb, al-Ṣawāʾiq al-ilāhiyya, 6, as quoted in, Bunzel, Manifest Enmity, 119.

46. iftarāʿ alā ahl al-ʿilm, Ibn ʿAfāliq, Risāla II, f. 54b, as quoted in, Bunzel, Manifest Enmity, 119.

47. waʾlladhī aqwaʾa hādhā ʿrajul fi hādhīhi ʿl-ʾarqāʾa al-ʾażīna annahu yanṣūru fi kutub Ibn al-Qayyim fiy fayā khudhu minhā mā wāfaqtu wa-yatruku mā khālafahu wa-yaʾkhudhu min awqāt wa-yatrakku akhirahu wa-nadhkuru lākum jumlat kalām Ibn al-Qayyim wa-shaykhihi ʾbn Taymiyya li-taʾlamī annaʾ bn ʿAbd al-Wahhāb dala wa-ʾadalla (“What brought this man into this terrible abyss is that he looks at the booksof Ibn al-Qayyim and takes from them what suits his fancy, disregarding what contradicts it; he takes from the beginning of a chapter and disregards the end of it. We will relate for you the entirety of Ibn al-Qayyim’s words and those of his teacher, Ibn Taymiyya, so that you know that Ibn ʿAbd al-Wahhāb has gone astray and led [others] astray.”), idem, Risāla I, ff. 45b-46a, as quoted in, Bunzel, Manifest Enmity, 119.

48. fa-yaʾ ibād Allāh hādhā ʿl-ṭāghūt fitana baʾd al-nās bi-kalām hādhayn al-shaykhayn fi awqāt al-mirḥāb yajīdu lahumā min al-kalām mā huwa madhkūr fi ʿl-Jahm ibn Ṣaţwān wa-Bishr al-ʿArabī wa-atbāʾ iḥimā min al-Jahmiyya wa ʾl-Muʾtazila wa-yaqraʾtu ʿaluʾ hāʾa ʿalā ilā ʿl-juhāl al-ʾawāmm ʾindāhu fa-yazumūna annahu yaʾnī al-sunna fa-fatanaḥum bihi ʾan dīnihiy fa-yuḥammilu kalāmāhum mā lā yaṭāmīlu, al-Bassām (“O servants of God, this ṭāghūt misled a number of people with the words of these two shykhīs early on. He would find words of theirs that were uttered in respect of Jahm ibn Ṣaţwān and Bishr al-ʿArabī, 87 and their followers from among the Jahmiyya and the Muʾtazila, reciting this to those ignorant commoners round about him so they would think that the Sunnis were intended by it. Thus he led them away from their religion, causing their words to carry a meaning that they do not bear.”), “Min asbāb al-muʾāraḍā,” 39 (transcription), 74 (manuscript photo), as quoted in, Bunzel, Manifest Enmity, 120.
three-letter root origins amounting to one who crosses limits are likely,⁴⁹ certain Orientalists have offered an alternative explanation by claiming it comes from the Ethiopic word *amlāka gēbt* (strange, foreign god), derived from the Greek *theos prosphatos*,⁵⁰ allegedly used by some to describe a false deity other than Allah.⁵¹ It is mentioned eight times in the Qur'ān⁵² and was used before the advent of Islam to refer to high deities in the pagan culture, including *al-Lāt* and *al-ʻUzzā*.⁵³ To date, it has been used to describe individuals who are considered apostates or are holding on to a dogmatic ideology that oppresses others.⁵⁴

The fact that Al-Razīnī employed this word to describe his scholarly opponent is reminiscent of the days of the critiques of Ibn Taymiyya. He seems to have failed to recognize that even Ibn Taymiyya did commit the heinous evil he casts Ibn ‘Abdu’l-Wahhāb as having done, which is decrying one’s opponents as Jahmiyya or Mu’tazila. To be clear, Ibn Taymiyya frequently considered his opponents to be the followers of Jahm ibn Ṣafwān and Bishr al-Marīsī, perhaps not truly so but as an epithet to discredit whoever’s views he did not favor. Both figures were controversial in the prevailing consensus of Sunnī Islam, with the Mu’tazila being sponsored by numerous Abbasid Caliphs and leading to clashes with many of the prominent scholars of their time. To date, many whose doctrines are inspired by Taymiyyan philosophy utilize the term Jahmī to cast other Sunnis as having gravely erred.⁵⁵ Yet Al-Razīnī’s view on the matter was not without opposition, as Ibn ‘Aflāqī was of the opinion that no deliberate misrepresentation had occurred. Hanbali scholars were keen to take this approach, as evidenced by Ibn Dāwūd essentially stating that Ibn ‘Abdu’l-Wahhāb was simply in another valley in relation to Ibn Taymiyya and Ibn al-Qayyim.⁵⁶ The independent reading exercised by Ibn ‘Abdu’l-Wahhāb was viewed by Ibn al-Amīr al-Ṣan‘ānī of Yemen as having opened the door to misreading and misrepresentations of the works of the earlier scholars⁵⁷ and that the lack of scholarly guidance to this young preacher led to him adopting whatever fit of Taymiyyan doctrines into his views despite its blind emulation being disliked.⁵⁸

---

⁵³. Fahd and Stewart, *Ṭāghūt*.
⁵⁵. Özervarli, *The Qur’anic Rational Theology of Ibn Taymiyya and His Criticism of the Mutakallimun*.
⁵⁸. wa-ghālib mā a’ma ‘ayn baṣiraṭthi wa-a’wa’iḥu fi zayghīhī wa-hayratīhī kutub Ibn Taymiyya wa bn alQayyim fa-inahu ‘ntahala l-muṣṭāla’ a fhā min ghayr ‘ilm wa-lā bāṣira wa-lā shaykh wa-lā dirāya
It is evidently clear to many early critics of Ibn ‘Abdu’l-Wahhāb that Ibn Taymiyya and Ibn al-Qayyim’s ideas have played a central role in shaping his thought. Taymiyyan ideology was placed on a pedestal, considered foolproof and almost as evidence itself. Muhammad ibn Hūmayd, the muftī of Mecca, who served during the 13th/19th century, claimed within his biographical dictionary of a Ḥanbali nature that Ibn ‘Abdu’l-Wahhāb was unwilling to even accept differing interpretations of texts produced by the aforementioned scholars if it contradicted his own doctrinal view.\

Ibn ‘Abdu’l-Wahhāb himself did not go to any lengths to hide his admiration for Ibn Taymiyya and his students. He went at length on praising them, claiming to not be calling people to follow a specific school of jurisprudence but that he looks up to the likes of Ibn al-Qayyim, al-Dhahabī, and Ibn Kathīr. In addition, Ibn ‘Abdu’l-Wahhāb praises Ibn Rajab for taking a steadfast position in condemning the practices of much of the masses. Their frequent condemnation of actions being ascribed as polytheism was looked favorably upon, with Ibn ‘Abdu’l-Wahhāb making frequent attempts to shore up his credibility by claiming to be following in the footsteps of some of the best scholars who arose in later generations.

It is definitely worth exploring the life of Taqī ad-Dīn ‘Aḥmad ibn ‘Abd al-Ḥalīm ibn ‘Abd al-Salām al-Numayrī al-Ḥarrānī, otherwise known as Ibn Taymiyya, who lived in the mid-13th century to early 14th century and arose as a jurist and theologian from the Ḥanbali school. He spent a large portion of his life in Damascus, with much of his teachings in the Levant that spread in the Islamic world being a root cause of controversy to his person during the Mamlūk Sultanate. He was subject to a lot of criticism and refutations in his lifetime, and his name and works have rose again in the contemporary era and are often cited in

---

59. “The thing that most blinded him and caused him to fall into his perversion and his confusion was the books of Ibn Taymiyya and Ibn al-Qayyim. He read them without knowledge or discernment, and without a teacher or clear understanding. Thus he would take from them what he imagined to be in accord with his fancies and leave aside what went against them.”), Ibn Dāwūd, al-Ṣawāʾiq wa l-ruʿād, f.35b, as quoted in, Bunzel, Manifest Enmity, 121.


61. al-mutaʾakhhirīn, sādatuhum wa-aʾimmatuḥum wa-aʿlamuhum wa-aʿbaduhum wa-azhaduhum, qad ishtadda nakīruhum ʿalā ahlʾ aṣsīrihim, Ibn Ghannām, Tārīkh, 1:248–49, as quoted in, Bunzel, Manifest Enmity, 122.

62. sādāt al-mutaʾakhhirīn wa-qaḍatatuhum, wa-kalāmuḥum fi inkār ḥādhā akthar min an yuḥṣara, Ibn Ghannām, Tārīkh, 1:446 (letter to Abd al-Wahhāb ibn Ḭūsā), as quoted in, Bunzel, Manifest Enmity, 122.
theological circles. He was widely respected for his immense knowledge that spanned different genres of the Islamic sciences, with many of his critics conceding to the near-prodigy status that he had obtained. A fervent engager in the polemics and refutation culture, Ibn Taymiyya authored numerous refutations of the Muʿtazila and Shiʿa, as well as other philosophers and theologians, particularly those who dabbled in kalām. Ibn Taymiyya received much backlash from the contemporary Sunnī scholars of the time, particularly those based in Egypt or Syria and of Ashʿarī and Shāfiʿī background. As a result, he spent many years battling prison sentences and court trials, wherein he absolved himself of any deviation by claiming to follow the Qurʾān, the sunna, and the words and deeds of the salaf, that is, the earliest generation of Muslims.

In some ways, it can be noted that Ibn Taymiyya was expanding on an already-developed base of Ḥanbalī thought, which was literalist and claimed to reject speculative theology (ʿilm al-kalām, or simply kalām), in direct opposition to the thought processes of much of the Ḥanafī, Shāfīʿī, and Mālikī scholars who had accepted Kullābī thought via Ashʿarism or the synthesized branch of Ḥanafī-Māturidism arising from Transoxiana. Opposition to the rationalism championed by the followers of Abū ʿI-Ḥasan al-Ashʿarī and Abū Manṣūr al-Māturīdī, known as mutakallimūn, was led by the those known as ahl al-hadith. This traditionalist and literalist approach was embodied within the Ḥanbalī school, who by and large stressed their following of only the Qurʾān, the sunna, and the salaf. Issues of creed were handled by blind taqlīd of those salaf that championed a literalist approach, with utter and unequivocal rejection of other forms of theological development or forms of thinking. The Ḥanbalī school, the smallest in comparison to the Ḥanafī, Shāfīʿī, and Mālikī schools, was considered synonymous with the traditionalist approach, with many recognizing its distinction as a school of law and theology. This was a largely unique feature for the Ḥanbalis, though some have contested the Ḥanafī school that enjoyed the position of being a school of theology as well.63 This was embodied in the person of Muhammad ibn ʿAbdallāh al-ʿAlawī, a Moroccan ruler who identified himself as Mālikī in matters of jurisprudence and as Ḥanbalī in matters of creed.64 Within his famous work of hadith collection, al-ʿAlawī goes on to describe Ḥanbalī theology as being pure, in that it refrains from indulging in the kalām within its paradigms and considers it to be consistent with the thinking of the major Sunnī imāms.65

---

63. See Rudolph, Al-Māturīdī and the Development of Sunnī Theology in Samarqand.
64. al-ʿAlawī, al-Futūḥāt al-ilāhiyya fi ahādīth khayr al-bariyya, 2nd ed. (Rabat: al-Maṭbaʿa al-Malakiyya, 1400/1980), 1, as quoted in, Bunzel, Manifest Enmity, 125.
outlook at the Ḥanbalī school, also known as Aṯharism, was viewed by many as being free of Judeo-Christian influences that tainted the development of Kullābī thought. In the eyes of Ibn Taymiyya, dwelling into kalām itself was something of a heresy. He frequently cited the remarks of prominent scholars of the four major Sunnī schools of jurisprudence to back his claims, including the Ḥanafī Abū Yūsuf by asserting that the founders all in essence agreed with his view.⁶⁶

It is worth analyzing Ibn Taymiyya’s claim regarding the Sunnī imāms and the positions of the founders of the schools of jurisprudence who have been lauded by a consensus of Muslims for their great works. Al-Shāfiʿī authored two books on the subject of theology, Tashīḥ al-Nubuwwah (The Validation of Prophecy) and al-Radd ʿalā al-Barāhimah (The Refutation of Brahmanism).⁶⁷ Despite engaging in debate with Ḥanafīs such as Bishr al-Mirīsī, he was reported to have adopted a harsher tone toward kalām. It is claimed he said, “If people knew the heretic tendencies kalām contains, they would flee from it as they do from a lion. It is better for a man to meet Allah with any sin save shirk than to meet Him with something of kalām.”⁶⁸ Imam Mālik bin Anas is reported to have been among the first to stipulate the principle of bilā kai⁶⁹ yet did not engage in persistently harsh condemnations of kalām as can be seen in the case of the eponymous founder of the Ḥanbalī school. Later scholars have attributed some of these positions as reflecting views on Muʿtazila thought, which remained the predominantly articulated example of kalām in Islamic theology, rather than on the science of it as a whole. In al-Baghdādī’s eyes, Imam Abū Ḥanīfa and al-Shāfiʿī were among the first jurists and founders of schools of jurisprudence to engage in theology.⁷⁰ Abū Ḥanīfa has a book attributed to him refuting the Qadarites, which is known as al-Fiqh al-Akbar. There is also the case of treatises, such as where he defended the view of the perceived Sunnīs at the time by declaring that one’s ability to act is formulated simultaneously with action itself, in opposition to the prevailing view among the Muʿtazila, which granted the capacity to conduct an action as existing before any tangible action is taken.⁷¹ It was, however, stipulated that this position could be perceived as being valid for two opposing ways, which would still fall

---

⁶⁶. man ʿalābā ʾl-dīn biʾl-kalām tazandaqa (“Whoso seeks [knowledge of] religion by means of kalām has become a heretic”), Ibn Taymiyya, Majmūʿ fatāwā, 16:473, as quoted in, Bunzel, Manifest Enmity, 126.
⁶⁷. Uṣūl, p. 308, as quoted in, Cerić, Roots of Synthetic Theology in Islam, 63.
⁶⁸. Ishārāt, p.36, as quoted in, Cerić, Roots of Synthetic Theology in Islam, 63.
⁶⁹. Literally translates to “without how.” It is usually used to mean “without asking how,” “without knowing how or what,” “without modality,” “without considering how and without comparison,” or “in a manner that suits His majesty and transcendence.” It essentially stipulates a non-committal approach to God’s attributes and essence and was used frequently in response to verses of the Qurʾān that literally translate to the face or the hand of God.
⁷⁰. Cerić, Roots of Synthetic Theology in Islam, 63.
⁷¹. Ibid. 63–64.
under the views of the established Sunnī community at the time. In the view of Qubāisah bin ʿUqbah, Imam Abū Ḥanīfa had engaged in theological disputes with heretics to the extent that he was a renowned expert in the field of such refutations until eventually choosing to step back from that science and position his focus in the matters of jurisprudence where he became further acclaimed in. In the case of Abū Ḥanīfa’s students, this dwelling into kalām only becomes significantly more evident, whose views cannot be discounted as they were integral to the shaping and early leadership of the Ḥanafi school. It is noted by Ibn al-Nadīm that a Khārijite by the name of al-Yamān bin Ribāb authored a book, titled Kitāb al-Radd ‘alā Hammād b. Abī Ḥanīfa (The Refutation of Hammād b. Abī Ḥanīfa), which showcased that Hammād definitely engaged in the kalām discussions to have an opponent critique his views. Al-Shaibānī is thought to have written a book titled ‘Aqāʾid al-Shaibānīyyah, which is found in the poetic format of Qaṣīdah Alfiyyah, which dwells into theological matters. Despite doubts having been raised with regard to exact authorship, given that commentators from the Shāfiʿī school seem to have engaged in this work a significant amount and that its report stems only from Ḥajjī Khalīfah, it is still evident this displays that al-Shaibānī did at the very least engage in issues pertaining to kalām, even if the precise book is a compilation of his views at a later date by another Sunnī scholar. In the case of al-Ḥasab b. Ziyād al-Lu’ūfī, one finds that he authored at least two books relating to the subject of kalām, Kitāb al-Maqālāt and Maʿānī al-Īmān (Meaning of Faith). Bishr b. Ghayyāth al-Mīrīsī spent a lot of time with Abū Ḥanīfa, and following his demise with his advanced student, Abū Yūsuf. Al-Mīrīsī was so involved in kalām that a distinct school known as al-Mīrīsīyyah formed out of his views. Despite being later rejected for its Murjiʿah tendencies, views on the Qurʾān’s createdness, and the position on faith being solely an internal matter of the heart, what it does reveal is that al-Mīrīsī along with other early scholars, including the imāms and those who studied under them, dwelled into theological matters and kalām.

Apart from Ibn Taymiyya’s extensive working in Islamic theology and utilizing an array of devices to defend his views, he also notably presented a new strain of thought within the Ḥanbalī approach that significantly altered the frame of thinking and responses to Ashʿarī thought, which was considered an innovation and a deviancy of sorts. Many of these alterations found a prominent place within Wahhābism. His writings are extensively filled with emphasis on establishing a rapport
between revelation (naql) and reason (‘aql). Some have attributed the Ash’arī Universal Principle (al-qānūn al-kullī), as postulated by Fakhr al-Dīn al-Rāzī, as constituting the notion that rational proofs must be given a higher weight in determining something in the event of a perception of conflict with proofs that have been revealed. In attempts to refute Ash’arīs, Ibn Taymiyya broke with a long-standing tradition in Ḥanbalī thought by engaging in reason-based arguments, especially when it came to the topic of God’s attributes. His bid to showcase that no conflict of any sort exists between revelation and reason can be seen in the work titled Darʾ taʿārud alʿaql waʾl-naql (“Averting Conflict between Reason and Revelation”). Ibn Taymiyya advances the notion that the fitra, which is relegated a great stature, enables one to have monotheistic inclinations and recognize God’s existence due to it being an innate sense within a person. With that train of thought, dwelling into kalām to prove the existence of God is considered unnecessary in Taymiyyan ideals. Several scholars have suggested that the break with Ḥanbalī tradition by dwelling into reason-like arguments was not well received by the contemporaries of Ibn Taymiyya from among the aforementioned school. At the very least, it did not reflect the earliest sentiments of Ḥanbalī scholarship as evidenced by the statements from a strong proponent of Ḥanbalī thought, one who engaged in a lot of refutations himself, ʿAbd al-Raḥmān ibn Rajab. Located in Damascus, he was from a period slightly after Ibn Taymiyya and can be viewed as scorning the tendency to engage in disputation or arguments on such delicate topics regarding God.

Ibn Taymiyya’s view on divine creation and its relationship with free will, in terms of viewing the former as encompassing of the position and conception of the latter, as well as being with distinct purpose is sharply different from the

78. Bunzel, Manifest Enmity, 127.
79. Fitra can be described as original disposition, natural constitution, or innate nature. See Encyclopedia of Islam 3, s.v. “Fitra” (Jon Hoover).
80. Hoover, Ibn Taymiyya’s Theodicy, 39–44, as quoted in, Bunzel, Manifest Enmity, 128.
82. al-taṣaddī li-radd kalām ahl al-bidaʾ bi-jins kalāmihim min al-aqyisa al-kalāmīyya wa-adiillat al-ʿuqūl . . . yakrahuhu ʾl-imām ʿAbd al-Raḥmān ibn Rajab al-Ḥanbali (”The Imām Aḥmad [ibn Ḥanbal] and the leaders of the ahl al-ḥadīth detested . . . refuting the innovators by partaking of their opponent’s discourse, that is, the use of kalām-like analogies and rational proofs . . . They deemed refutation appropriate only by the texts of the Qurʾān and the sunna, and by the words of the pious ancestors, if such were to be found. Otherwise they deemed silence to be preferable.”), ed. ʿAṯʿalāʾ, Taʾlīḥ al-farrāʾ al-ḥulwānī, 2nd ed., 5 vols. (Cairo: Dār al-Ḏāriʿal-Ḥadīth vol. 1434/2012), 2:637–38; translation borrowed from Caterina Bori, “Ibn Taymiyya wa-Jamāʿatu-hu: Authority, Conflict and Consensus in Ibn Taymiyya’s Circle,” in Ibn Taymiyya and His Times, 23–52, at 36, with minor changes, as quoted in, Bunzel, Manifest Enmity, 128.
Ashʿarī or Ḥanbalī positions articulated previously. Ashʿarism remained a strong proponent of the doctrine of voluntarism and emphatically rejected the notion that God creates with a specific cause (ʿilla) or purpose. It was stipulated that such an underlying purpose or cause formulates the implication that such a purpose or cause would need to have existed before it engages in subsisting in God. In a bid to downplay the deterministic attitudes displayed by the proposition that God creates all good and bad things, which encompasses human acts (afʿāl), the Ashʿarīs forwarded the doctrine of acquisition (kasb) wherein human beings would engage in acquiring acts, regardless of them being good or bad, immediately prior to enacting them. In this way, God’s omnipotence was affirmed while being able to comfortably reject a needed purpose to God’s acts. On the contrary, the Ḥanbalīs at the time were known to have engaged in a literalist affirmation of passages of the Qurʾān that dealt with matters at hand while abstaining from any inquisitive outlooks into their meanings. Matters pertaining to creation and predestination are encapsulated in statements such as “He guides whom He wills” (yahdī man yashāʾu). Ibn Taymiyya articulated a vastly different outlook on the subject, affirming that God does indeed act on behalf of a purpose that befits his wisdom (ḥikma), which, in turn, would render people to be responsible for their individual acts. In some accounts, it is viewed that Ibn Taymiyya and Ibn al-Qayyim were of the view that hellfire could not be eternal, in contrast to paradise, as it would not fit the wisdom of God. Naturally, this train of thinking necessitated the use for independent interpretation of what would befit the wisdom of God.

Conclusion

As one looks at the legacy of Muhammad ibn ʿAbdu’l-Wahhāb, it is difficult to not draw parallels with the works of Ibn Taymiyya. Perhaps what sets apart this strand of thought the most was its unapologetic approach to accepting elements of Taymiyyan thought that were different from those found in the Ḥanbalī tradition prior to that. The 14th-century scholar’s views on theology, jurisprudential issues ranging from triple talaq to the state’s role in governance, and the visitation of the

---


tombsoftheprophetsandthesaints (ziyarah) set him in stark contrast to much of the rest of existing Sunnī scholarship. Contemporary Wahhābism continues to undergo shifts and changes yet has begun formulating the premises of a system of taqlid (imitation) of scholarship dating back to Ibn ‘Abdu’l-Wahhāb and, at times, Ibn Taymiyya. Shāfiʿī and Ḥanbalī critics’ differences on Wahhābism’s classification in relationship to Sunnism are one that seem to rest on the extent to which parallels are drawn to Ibn Taymiyya and the defining parameters of orthodoxy as it relates to theology. It is not unprecedented for a movement to be theoretically misaligned with Sunnī thought while still adhering to some jurisprudential principles and acceptance of the Four Caliphs, as seen in the case of the Muʿtazila. Further research needs to be conducted to look at contemporary and historical definitions of Sunnīsm and what matters of jurisprudence and theology are vital components for inclusion or exclusion. While it may be unclear whether Wahhābism could be considered a madhhab, it is undoubtedly visible for many Sunnī critics that it is a strand of Islam that needs to be recognized as heterodox in nature.

The exploration of Wahhābism and its Taymiyyan origins is not merely an academic exercise; it holds profound significance for both Islamic and world history. The rise of Wahhābism has had lasting impacts on religious, political, and social landscapes, shaping ideologies and influencing relations among various Islamic communities and the wider world. Understanding its historical roots provides crucial insights into contemporary issues and conflicts, helping to navigate the complex interplay of tradition, reform, and modernity. By examining this phenomenon, we can better appreciate the multifaceted nature of Islamic thought and its role in global affairs, answering the vital question: why should we care? The study thus transcends the boundaries of historical scholarship, becoming a touchstone for policymakers, scholars, and anyone interested in the interconnectedness of religion and global dynamics.

References


Qamar, Zubair. 2012. “Wahhabism, Understanding the Roots and Role Models of Islamic Extremism.”


In a growing society, healthcare has become an essential part of a country’s development, standard of living, and life expectancy. The past decade has seen unprecedented growth in both scientific and medical discoveries, leading to improved treatment plans and innovative treatments, yet there are some aspects of healthcare that remain stunted. Discrimination, particularly on the basis of race, is still prevalent and ingrained in the American healthcare system, leading to worsened patient outcomes for people of color and a discomfort with seeking medical care that disproportionately affects those in marginalized groups. This bias continues even into the education of healthcare providers, which perpetuates the cycle of prejudice in the medical field.

While the intersection of racism in the medical field varies greatly and maintains its prevalence among different races, this dissertation focuses mainly on discrimination of black patients in the American healthcare system. For further information on the effects of racial bias, both abroad and among a wider variety of races, valuable resources can be found below the works cited in this review.

The sources analyzed in this review are analytical and scientific, with little to no use of primary anecdotal evidence. While such an emotional topic as racism and how it affects patient care relies heavily on personal experience, these sources have confounded the feelings of many patients nationally and adapted them into a well-reviewed analysis of how those feelings reflect nationwide opinions. Furthermore, it includes historical papers detailing the foundations of the medical field and how it was developed in hand with racism.

The following sections include definitions of terminology surrounding the issue, information on the history that led to modern-day biases, the ways in which these biases are seen by patients, and how it affects their care. This demonstrates the link between past practices and the discriminatory actions in the medical system currently. Furthermore, it establishes evidence that there are effects felt by patients of color as well as offers information on possible solutions to help better their treatment.
terminology

This review serves as an inquiry into the literature surrounding the attitudes, research, and effects of racial bias in healthcare settings. It is first necessary to delineate the scope of this review by defining some key terms and points of interest for this paper. These include diversity, discrimination, including macro- and microaggressions, and different types of racism. To begin, diversity refers to the inclusion of healthcare professionals, trainees, educators, researchers, and patients of varied race, ethnicity, gender, disability, social class, socioeconomic status, sexual orientation, gender identity, primary spoken language, and geographic region (1). Moreover, discrimination, specifically in healthcare, refers to negative actions or a lack of consideration being given to an individual or a group that occurs because of a preconceived and unjustified opinion (1). Individuals do not necessarily need to be members of these groups to experience discrimination against that group, if they have perceived membership by the oppressor. Dictionary.com defines oppressor as “a person or group that exercises authority or power over another in a harsh and burdensome way.” In this way, the oppressor must be the person who benefits from the system of prejudice in place, which in this case may be white or white-presenting people. Some common reasons for discrimination toward an individual may be because of the individual’s race, ethnicity, gender, disability, social class, socioeconomic status, sexual orientation, gender identity, primary spoken language, or location of residence. There is an inverse relationship between discrimination in healthcare and diversity in healthcare, meaning that the more diverse the medical staff is, the greater the quality of care for patients in marginalized groups. This review will be mainly focusing on the effects of discrimination to patients of color, as opposed to the higher barrier of entry for people of color seeking a profession in the healthcare field. However, as evidence shows that discrimination of marginalized patients is lowered with higher rates of diversity in the healthcare field, it is clearly important to consider that aspect as well (1). While it considers the tangential effects of the workers’ racial backgrounds, this paper focuses predominantly on the clinical ramifications of discrimination in the medical field on patients receiving care.

Within discrimination, there are two main classifications of actions: macroaggressions and microaggressions. Microaggressions are overt and more radical forms of racism that are rooted within society or within the medical system. An example would be theoretical laws that require testing of potential medical treatments on white patients but not those of color, preventing the full effects of medicine from being seen on a wider range of patients. Macroaggressions
have been largely eliminated compared to decades earlier in history, with laws such as Title VII of the Civil Rights Act and the passing of the Americans with Disabilities Act in 1990, which prohibit unequal treatment based on race, sex, and disability (1). While protective legislation such as this has decreased the overt discrimination associated with macroaggressions, microaggressions have increased in response. Microaggressions are short, everyday insults or remarks that can be difficult to identify due to their barely perceptible nature. Nonetheless, they convey a negative message to a person because of their affiliation with a marginalized group. Due to the nature of microaggressions, they tend to be easier to perform, harder to identify, and subsequently more difficult to punish.

Microaggressions have been associated with damaging the victim’s mental health through summaries of nationwide surveys and interviews of patients. The takeaway from these interviews is that microaggressions tend to cause lower self-esteem, worsened self-care, increased susceptibility to substance abuse, depression, suicidal ideations, anxiety, and more (1). Recent studies have even shown that regular exposure to microaggressions is associated with a higher incidence of hypertension, increased frequency of hospital admission, and more severe diabetes-related stress (1). As microaggressions are delivered predominantly to patients of color, this results in marginalized groups of people receiving a disproportionately negative healthcare experience and adverse effects on their health. Furthermore, while macroaggressions are much more overt, seen predominantly in legislature and other tangible aspects of society, microaggressions are typically delivered in one-on-one scenarios, making them even more difficult to identify and stop. The main differentiator of microaggressions is that they are more social in nature and thus harder to provide evidence for their occurrence.

Furthermore, according to the National Museum of African American History and Culture, there is individual racism, interpersonal racism, institutional racism, and structural racism (3). Individual racism is most directly related to the biases which we hold or rather our personal beliefs in the superiority of one’s race over another. Interpersonal racism is an expression of these biases between individuals, institutional racism is shown in the policies and procedures of an organization, and structural racism is the total effect of these agents across systems and between institutions. These are all forms of racism that build a foundation for the discrimination being discussed in this review, and in order to promote health equity across races, it is required to address both individual and interpersonal racism while dismantling the institutional and structural racism that is built into the crevices of our society. Further examples and definitions from the National Museum of African American History and Culture are summarized in the next table (3).
Literature Review: The Roots and Clinical Effects of Racial Bias in Medicine

<table>
<thead>
<tr>
<th>Term</th>
<th>Individual Racism</th>
<th>Interpersonal Racism</th>
<th>Institutional Racism</th>
<th>Structural Racism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>“The beliefs, attitudes, and actions of individuals that support or perpetuate racism in conscious and unconscious ways” (3).</td>
<td>Occurs between individuals, including public displays of racism, such as slurs or prejudicial actions.</td>
<td>Occurs in organizations, seen in the form of race-based policies and practices that offer an unfair advantage to white people over Person of Color (POC).</td>
<td>“The overarching system of racial bias across institutions and society” (3).</td>
</tr>
<tr>
<td>Example</td>
<td>A white person tells a racist joke.</td>
<td>A white person uses a racial slur against a POC.</td>
<td>Schools with the highest percentages of students of color tend to have the least funding.</td>
<td>Pop culture portrays POC as criminals by depicting them as such in roles.</td>
</tr>
</tbody>
</table>

Lack of African American Physicians

Although this paper focuses more on the clinical ramifications of racial bias toward patients of color, the reasons the majority of hospital staff are white must also be addressed. Much of the racism rooted in today’s society is remaining from decades of discrimination in the United States on a greater legislative basis. This means institutional racism in the form of governmental practices, which inadvertently place POC at a disadvantage to white citizens. One example of this is schools’ dependence on local property taxes, which provides affluent (more commonly majority white) communities with greater access to education than impoverished communities, which house majority POC. This can be traced back to the dawn of slavery, preventing African Americans from garnering the same education as their white counterparts; however, this review will begin with the Flexner Report in 1910. Abraham Flexner was a member of the Hopkins Circle, a group created to place a foundation of science-based medical training in the United States. He evaluated medical schools in the United States and Canada from the point of view of a teacher and subsequently destroyed the reputation
and potential funding opportunities to those schools whose education he deemed inadequate (4). Among the majority of these schools were medical schools for African American students, and the community is still recovering from the result of this report. This report significantly slowed the inclusion of African American physicians in the American medical system while creating a narrative that African American physicians are less qualified and able to succeed than their white counterparts (4).

Concordant care is a patient sharing a common attribute, such as race, gender, or ethnicity, with their provider. Evidence strongly supports the conclusion that race-concordant patient-physician relationships correlate with improved communication, longer patient visits, greater adherence to medical protocols, and higher patient satisfaction scores (1). Therefore, the lasting effects of racism in the United States from things such as the aforementioned Fletcher report show a clear effect on patients of color, as they are not able to receive this race-concordant care that has proven positive effects on a frequent basis. Even more, underrepresented minority physicians are more likely to serve in areas with a physician shortage and serve underserved populations, such as those in low-income areas and minorities (1). With fewer minority physicians, there is less outreach to these individuals; thus, the effect of racism is clear through its ramifications on patients.

### Historical Racism in Medicine

As with much of the racism that is ingrained in American culture, the roots of medical racial bias can be traced to slavery, most notably with the perfection of the surgical technique for vesicovaginal fistula, or VVF. VVF affected many women in the 19th century, resulting from obstructed labor that caused a tear from the bladder to the vagina (5). This left many women incontinent with a continuous leak of urine, forcing many victims to social outcasting and, later, suicide. The field of gynecology did not exist in the early 1800s when this issue was most prominent, and the examination of female organs was considered disgusting for doctors, who were majority men at the time. In performing pelvic examinations, doctors looked women in the eyes, not even being able to acceptably look at the vagina during their evaluation (5). Even in medical schools, obstetrics was taught with dummies, and doctors generally did not see live birth until they were in charge of delivering a baby themselves (5).

Dr. J Marion Sims, the American surgeon deemed the father of gynecology, perfected the first usable surgical technique of VVF in 1849, after forcing enslaved women for four years to undergo experimental surgeries. Using a speculum made from a pewter spoon, at age 27, he used a total of seven enslaved
women with VVF in his experiments, all without anesthetics, as Sims was not aware of the developments made in this area of medicine (5). The first woman was Lucy, who underwent an hour-long invasive operation with 12 local doctors gawking at her. The operation failed, and she became ill with fever from blood poisoning, recovering after a matter of two to three months. A second victim, Anarcha, was operated on a total of 13 times until a cure was found. After this, all of the enslaved women’s conditions were corrected, and they were sent back to where they lived. It is important to note that all of these operations were done with the permission of the “owner” and not the patient, and when white women came to Sims for the procedure on their own accord, not a single one was able to endure the pain and finish the operation (5).

Even today, there is a medical school named after Sims and statues celebrating him in New York and South Carolina, with many deeming him the “father of gynecology.” It was normal in this time period to force enslaved peoples to undergo experimental procedures, and it was valuable in order to find a cure for the condition. However, significant medical breakthroughs were being made in this period without the exploitation of enslaved people, and Sim’s use of involuntary procedures was not a common practice (5). To this day, the stigmatization of female anatomy is present in the medical field, with complete vaginal anatomy not being taught in textbooks. While medical students are commonly taught about the pleasure sensors of male genitalia and how to treat issues associated with them, such as erectile dysfunction with Viagra, there is little focus on clitoral anatomy and how female pleasure centers operate. This leaves a large gap in knowledge that can contribute to incidents of clitoral atrophy and other related illnesses being left untreated. Furthermore, a 2013 study published in the *Journal of Obstetrics and Gynaecology* examined 59 gynecology and anatomy textbooks for information on the dimensions of vulval constituent parts, and none of them gave measurements for all vulval structures (6). Of those that contained measurements for one or some of them, the ranges were much narrower than recent studies suggest. The result of this knowledge gap is exemplified in a survey of 433 Australian general practitioners, wherein only 75% said they were confident in assessing genital appearance, which is a basic part of women’s health (7).

Subsequently, this history of forced procedures has continued into a stereotype that black people (especially women) are not as susceptible to pain. This leaves the reader with the questions of how black people, and women, can trust a system that praises a man who tortured women and trust providers with a skewed image of their anatomy and physiology. With this, it is important to note the intersectionality between gender and race in medical bias, as women of color have a multitude of stereotypes and issues to deal with. However, the scope of this paper does not include a lengthy analysis of this intersection, and for further
discussion on the topic of gender bias in the medical field, one should read the directed papers following the works cited.

Yet another example of historical discrimination comes from the Tuskegee airmen, a group of primarily African American military pilots and airmen who fought in World War II. In 1932, the United States Public Health Service (USPHS) in hand with the Tuskegee Institute began a study recording the natural history and prevalence of the STD syphilis (8). Involving 600 black men, 399 with syphilis and 201 without, researchers told the participants that they were being treated for “bad blood” in exchange for free meals and burial insurance. This term was used to describe a variety of illnesses, from syphilis to anemia and fatigue, so the participants’ informed consent on the project was not collected, as the true nature of the study was not disclosed (8). By 1943, penicillin was used as an effective treatment of syphilis and was widely available, yet participants were neither told about this treatment nor offered it, allowing the patients to suffer from a disease that had an effective and available cure. In 1972, this study was exposed, and the Assistant Secretary for Health and Scientific Affairs concluded that it was “ethically unjustified,” thus ending the study in October 1972, more than 29 years after the discovery of a cure (8).

Since then, the Tuskegee Health Benefit Program was established to ensure medical accessibility for affected participants and their families, which continues to this day. In 1974, the participants won $10 million in a class-action lawsuit, and a Presidential Apology was issued by Bill Clinton in 1977 (8). Despite these reparations, hundreds of black men suffered at the hands of induced ignorance by their doctors, something that never would have been done to white people at the time, who also suffered from syphilis but were never included in the study. The history of mistreatment of African Americans is rooted deeply in American medical culture and contributes greatly to a sense of distrust of black people in the medical system. This has many ramifications in how they obtain their medical care and their treatment, which is detailed in the following section describing the lasting results of this discrimination in today’s world.

**Modern-Day Clinical Ramifications**

Physicians holding an implicit bias against minorities have proven to create detrimental effects in their care. A recent study in *Proceedings of the National Academy of Sciences* of the United States demonstrates the connection between incorrect beliefs of biological differences between races and racial bias in both pain assessment and treatment recommendations (9). Medical students and residents in the study agreed that African Americans’ nerve endings are less sensitive to pain than that of white people, and their skin thicker, despite these facts being
unfounded. The medical professionals who incorrectly made these assumptions rate African American patients’ pain lower than that of white people, therefore resulting in less accurate treatment recommendations.

Yet another even more concrete example is in glomerular filtration rate measurements, used to measure how much blood biological filters in one’s kidneys clean every minute based on their body size. The so-called race corrected estimated glomerular filtration rate measurements are based on the unscientifically supported belief that African Americans have higher creatinine levels and more muscle (10). These facts being unsupported may result in a higher reported estimated glomerular filtration rate, which is interpreted as being the healthier renal function for African Americans, but in reality it is not. This can lead to a patient not identifying a possibly harmful renal issue and being falsely led to believe they are healthy, resulting in delayed treatment. Subsequently, in a medical artificial intelligence program, which considered past healthcare costs in predicting the clinical risk of certain patients, a larger majority of white people had greater rates of spending on healthcare and thus were determined to be higher-risk patients than African Americans (10). This leads to a possible underreporting of healthcare risks faced by the African American community.

Although bias is present in all areas of healthcare, one specialty it is especially dominant in is dermatology. There is an underrepresentation of darker skin tones in dermatologic texts, general medical texts, and scientific literature. This compromises the clinical tools of trainees with patients of color, as most of dermatology is identifying dermatologic issues by how they look on the skin. A 2006 study found that coverage of dark skin in images in major dermatology resources ranged from 4% to 18%, and it was excused as “harder to capture” in images (11). However, a 2020 study found the same thing, with up to 18% of images containing darker skin tones, even with the advancements in technology (11).

Within textbooks, this bias runs deeper with the associations being made between certain diseases and race. White skin is presented with more common skin conditions, such as acne and eczema, whereas darker skin is used overwhelmingly to show sexually transmitted infections. This demonstrates an implicit bias in the image selections of publications, which can be translated into the doctors learning from them and, subsequently, their patients. Even in non-dermatological fields, a 2018 study of general medical texts found that under 5% of images included dark skin tones, and only 18% of images in the New England Journal of Medicine included non-white skin tones from 1992 to 2017 (11). With the COVID-19 pandemic, many more medical students are relying on pictures as opposed to more hands-on training, and this will lead to even more of a bias toward lighter skin in treatment. This creates a hesitancy in clinicians to
diagnose darker skin tones and thus a compromise of patient care for patients of color, which can be seen greatly in the treatment of melanoma.

Although people with more melanin in their skin develop melanoma at a 20–30 fold lower incidence than non-Hispanic whites, it represents one of the largest disparities in survival for any cancer (12). Data from the National Center for Health Statistics suggests that for every three black men or women diagnosed with melanoma in the United States, one dies of the disease, whereas for non-Hispanic white men and women with melanoma, roughly one in seven and one in 11 die from the disease (12). A study conducted in 2006 calculated a two- to threefold greater risk of mortality among black patients with melanoma, and a 2016 study found similar results, pointing to “a serious disparity . . . in melanoma diagnosis and outcome for white patients compared with minorities” (12). Moreover, initial studies conducted establishing a strong association between UV exposure and melanoma completely excluded participants with darker skin types (12). Even with a correlation between increased melanin and sun protection, acral lentiginous melanoma (ALM), which is not affected by sun, makes up a proportionately higher percent of melanoma cases in darker-skinned individuals. Thus, the Eurocentric emphasis on solely wearing sunscreen may overshadow ALM risk, as well as lead people of color into an incorrect assumption that they are free of melanoma risk and do not need to seek treatment, leading to later diagnoses (12).

These disparities do not just relate to melanoma, but the American Cancer Society’s Cancer Facts and Figures for African Americans 2019–2021 concluded that black patients face a survival gap for most cancers, which results much less from biological differences than from socioeconomic and racial disparities that result in unequal access to work, income, education, housing, healthy food, high-quality healthcare, and an overall unequal standard of living (12). With this, black patients are significantly more likely to be diagnosed with later-stage melanomas than their white counterparts (12). It is clear that racial bias in medicine stems from the very basis of medical students’ education into affecting black patients’ mental and physical health. Substantial evidence that supports this conclusion is found in study after study and points to the necessity for immediate work toward a solution.

Solutions

While the issues of racism in healthcare are rooted in the implicit, subconscious bias of every person involved as well as in the history of the industry, there are ways to improve these effects on patients of color. One skill for healthcare professionals to develop is focusing on the person behind the patient. Instead of relying on
preconceived notions based on factors such as race, the provider should foster a communication with the patient wherein they can learn about them as a person, their behaviors, and other factors that may affect their care. It is also beneficial for physicians to undergo training to recognize their own implicit biases so that they are able to pinpoint and stop them as they happen. However, providing professionals with too much race-based training may lead to strengthening their stigmatization of certain characteristics without promoting healthcare outcome improvements. It is important to balance recognizing differences associated with race with the idea of patients as multifaceted identities. The objective is not to be color-blind, but instead to not be blinded by color, and acknowledge its effects in hand with other aspects of the patient’s identity.

Furthermore, it is essential to teach future students to provide equal care to all races, starting with including a wider variety of images of people with darker skin tones in medical textbooks. Maline Mukwende, a medical student at St. George’s University of London, worked with the school to create a guide titled Mind the Gap, which compared images of cutaneous and systemic diseases side by side on both dark and light skin tones (Kaundinya Kundu). On the patient side, the Skin of Color Society works on recruiting, retaining, and training more board-certified dermatologists of color. They also sponsor “Find a Doctor,” which is aimed at improving patients’ access to board-certified dermatologists who specialize in skin color (12). This starts from the source and will lead to future physicians being more comfortable and familiar with diagnosing patients of color, something found to be essential to their comfort and even their mortality.

References


11. Kaundinya T, Kundu RV. Diversity of Skin Images in Medical Texts: Recommendations for Student Advocacy in Medical Education. Journal of Medical Education and Curricular Development. 11 Jun 2021;8, doi:10.1177/23821205211025055

Extensive Drug-Resistant Typhoid Fever Prevention and Management in Pakistan: A Challenge to Public Health

BILAL IRFAN, IHSAAN YASIN, AND DENISE KIRSCHNER

Typhoid fever is a bacterial infection that largely spreads through contamination in food and water, as well as by close contact, and displays many cold-like symptoms in addition to more severe gastrointestinal, muscle, and life-threatening states. The emergence of a new, extensive drug-resistant (XDR) strain in Hyderabad, Pakistan, in 2016 resulted in the need for stronger antimicrobials to combat the pathogen. Despite vaccination and contact tracing interventions, the multi-drug resistance and fitness of the H58 strain Salmonella enterica serovar Typhi contributed to the rapid spread of typhoid fever in Pakistan. Vaccination is at the forefront of efforts attempting to combat XDR typhoid fever cases in Pakistan, and public education systems and schools should prioritize health classes that pertain to the spread of the disease and what steps individuals can take to be safe, as well as social media infographics circulated through Instagram and Facebook with popular hashtags to reach target audiences. A robust and forward-thinking approach needs to be taken by the Pakistani government to use its budget and international funding in the sectors of health and education to ensure disease prevention.

Introduction

Typhoid fever, a result of the Salmonella enterica serovar Typhi (S. typhi), is a bacterial infection that largely spreads through contamination in food and water, as well as by close contact. Those infected by S. typhi may develop typhoid fever (typhoid), displaying many cold-like symptoms in addition to more severe gastrointestinal, muscle, and life-threatening states. Fevers of up to 103–104°F (39–40°C) are commonplace alongside coughs, stomachaches, and lethargy.1,2 South Asia is a leading region in the world in the number of typhoid cases, with

Contact: Bilal Irfan <birfan@umich.edu>

https://doi.org/10.3998/umurj.5505
a significant portion of populations at risk, including immunocompromised adults and children. Pakistan has the highest incidence of typhoid among its neighboring countries, averaging 493.5 cases per 100,000 in 2018. The country suffers from issues of overcrowding, inadequate sanitation and healthcare, and insufficient access to clean water, ranking at 35.84% availability of clean water in 2020. Deteriorating living conditions enable food and water sources to become contaminated with feces that contain bacteria, with poor water, sanitation, and hygiene infrastructure contributing to the spread of the illness.

The virulence and spread of S. typhi have been influenced by its ability to endure for long periods of time outside of a human host, such as in water or on surfaces, as well as other genetic factors. In 2016, the emergence of a new extensive drug-resistant (XDR) strain of typhoid in Hyderabad known as H58 resulted in the need for stronger antimicrobials to combat the pathogen. The advent of COVID-19 exacerbated the public health crisis because overlapping symptoms made identification of case etiology difficult, as well as explorations of treatment options for COVID-19, including azithromycin, which could inadvertently increase resistance to typhoid across the scope.

In this article, we conduct a comprehensive review of the literature on the state of XDR-typhoid in the post-COVID-19 era, develop insights on the conditions that promoted the spread of the disease, and highlight recommendations to reduce the burden of the disease. We look through databases, including PubMed and Google Scholar, to find relevant research articles, academic journals, and authoritative reports. This was filtered down to those pertinent to the study’s geographical reason, with the review serving as the foundation of knowledge regarding the state of XDR-typhoid and its implications for global health. In addition, data collection was a crucial part of the study, with reliable and up-to-date information on XDR-typhoid cases, epidemiological trends, and public health interventions gathered from reputable sources such as government health agencies, international organizations, and research institutions. The government of Pakistan’s own reporting of the epidemic and national trends through its weekly field epidemiological reports was particularly useful. The collected data was then analyzed and synthesized to identify key trends, patterns, and insights, shedding light on the factors that contributed to the spread of XDR-typhoid and the challenges faced in managing the disease. Furthermore, the research aimed to identify any knowledge gaps or areas that required further investigation, guiding future research endeavors.

Socioeconomic Context and Water, Sanitation, and Hygiene in Pakistan

Issues ranging from contaminated water being used in drinking and irrigation sources, as well as the overall weak socioeconomic status of many Pakistanis,
exacerbate the spread of typhoid. In Pakistan, economic turmoil and foreign debts, which leave just $151 million for the country’s whole healthcare system, are important contributors to the outbreak of typhoid. Given that over 40% of the population that falls below the poverty line reside in slums and congested regions, the current rates of investment in healthcare fall acutely short of fulfilling the country’s needs. Airborne infections are caused by inadequate sanitation facilities, lack of clean food supply, and inaccessibility of clean drinking water. In addition, insufficient money prevents the provision of acceptable residential infrastructure and common sanitary practices despite the development of basic health services established by government agencies in neighboring municipalities. Thus, a lot of doctors hesitate to work in rural regions. As a result, many people are unable to access medical care or guidance, increasing the incidence and mortality of typhoid.

Potential carriers from impoverished regions go to urban centers in pursuit of secure employment or as a last resort to escape poverty, leading to an increase in typhoid fever cases among city residents. Although plenty of individuals prefer food from the streets, there are significant rates of typhoid infections that are brought on as a result. Estimates range from 70% to 90% in reports of seeing animals, flies, other insects, or liquid waste in locations where meals are prepared. Food goods are frequently contaminated by dust and animal droppings when carts are parked near or over drains. Drains can be a source of water and food remnants, thus attracting insects, rodents, birds, as well as pests and other contaminants if not cleaned properly. There is also the potential for airborne transmission through windblown particles of dust and debris entering food products. Many vendors handle food without following standard sanitary procedures, such as cleaning their hands with soap and water after handling raw ingredients, cash, or food packaging. Cross-contamination is common since improperly cleaned surfaces can still have food preparation residue on them.

Most rural communities do not have access to services like faucets, wash basins, and individual hygiene supplies; thus, rivers and water wells are the main sources of drinking water there. In addition, a lot of the water they receive is not fit for human consumption because it has not undergone quality testing and may have high bacterial loads. Furthermore, companies that use their property as a landfill pollute their water supplies. All of these elements work together to severely pollute the environment and contaminate their sole water supplies, raising the possibility of developing Salmonella infections. Due to poor personal as well as sanitary hygiene practices, rural communities are more susceptible to typhoid. This is clearly given in that over 25 million individuals in Pakistan practice open defecation in vegetation, waterways, and on highways, further increasing the potential for pathogen transmission.

Karachi has been noted to also have the presence of many microbial contaminants within its various water supplies, suggesting how the consumption of
water by the 16 million residents or so of the city can be a means of advancing the spread of the disease.\textsuperscript{20} The unhealthy effects of it are notable, and the increase in population in the city center as a result of migration from rural outskirts to urban areas has led to predictions of the population projecting to be 23.1 million strong by the year 2035.\textsuperscript{21} A significant section of the country’s population is now forced to reside in impoverished and congested places, where access to necessities like a functioning sewage system, drinkable water, and clean food is simply out of reach. To make matters worse, residents in these places are less informed about fundamental hygiene principles and the transmission of infectious illnesses, which leaves them more open to the dangers of typhoid fever. In addition, these communities’ residents lack simple access to institutions that provide economical and high-quality healthcare. Furthermore, the existing antibiotic resistance is getting worse due to a shortage of qualified medical staff working in such institutions.

\textit{Spread of Typhoid in Pakistan}

The end of 2016 saw the advent of the first recorded case of an XDR strain of typhoid as it emerged from Sindh, Pakistan. Initial reports placed it in conjunction with the massive influx of typhoid fever cases that had been confirmed via blood culture and acted refractory to standard therapy.\textsuperscript{6} From 2016 to 2021, the National Institute of Health (NIH) Islamabad Weekly Field Epidemiological Report reported a total of 14,360 XDR-typhoid cases in Karachi alone, the largest city and capital of the Sindh province. During a similar period spanning from November 2016 to June 2021, a total of 5,741 confirmed XDR-typhoid cases were reported in all the remaining districts of Sindh, discounting Karachi. Of those, 69.5\% of cases came from the Hyderabad District.\textsuperscript{7} As a result of this, Pakistan began spearheading the implementation of the typhoid conjugate vaccine as a part of its routine childhood immunization program, as per the guidelines stipulated by the World Health Organization (WHO).\textsuperscript{6}

Fear of international outbreaks and spillovers from the currently multidrug-resistant strain of typhoid fever in Pakistan remains at the forefront of the concern of many global actors.

The onset of the COVID-19 pandemic drastically altered the landscape of the discourse around typhoid fever as it began to pose possibilities of co-epidemics and co-infections.\textsuperscript{17, 18} The healthcare infrastructure in Pakistan has been noted to be already burdened beyond its capacity, and several pulmonological conditions including acute respiratory distress syndrome, pneumothorax, or even the advent of other cardiovascular diseases and secondary infections can drive the fragile system into collapse.\textsuperscript{19} Such scenarios came to a head when 20,000
typhoid cases were reported to have occurred within a span of 10 days in the summer of 2020 with co-infections of SARS-CoV-2.5

Healthcare practitioners all around the country are finding it difficult to control the XDR-typhoid strain given the variety of challenges faced. Azithromycin, an antibiotic used to treat a wide range of bacterial infections, is still the sole accessible oral treatment choice among the three medicines frequently used to treat XDR-typhoid fever, making it the final line of defense for treating patients in an outpatient environment. Nevertheless, XDR-typhoid is becoming more resistant to the administration of azithromycin, according to recent investigations.22 As a result of this, physicians are being forced to inject both tigecycline and carbapenem, which are typically reserved as the last-resort treatment for typhoid fever. Due to their high cost and relative inaccessibility to the impoverished population in developing nations, many of these medications require in-patient treatment.11 In addition, when instances of XDR-typhoid fever epidemics started to climb at an alarming rate, authorities in a number of nations expressed worry over the coming global impact of these outbreaks.

**Figure 1:** Extensive Drug-Resistant Typhoid Cases per Year in Sindh Province, Pakistan (2017–2021) (This graph was formulated using data released in the Pakistan Weekly Field Epidemiology Report (08/25/2021) by the National Institute of Health (NIH) Islamabad. Data for 2021 is only until August 21.)
Travel restrictions and guidelines are already becoming commonplace as seen by CDC recommendations urging all travelers to this region of the world to have a typhoid vaccination as a precaution.\textsuperscript{23}

\textbf{Multi-drug-resistant Strain of \textit{S. typhimurium}—Haplotype 58}

Typhoid fever is primarily caused by the bacterium, \textit{Salmonella enterica} serovar \textit{Typhi}, which is able to spread easily due to a variety of factors. The presence of the typhoid toxin that can bind to a multitude of cells and its possession of a polysaccharide capsule aids its ability to rebuff the immune system and survive in harsh environments, such as an acidic stomach. \textit{S. typhi} demonstrates the capacity to generate virulence factors, encompassing toxins and enzymes, facilitating its intracellular invasion and replication within human cells.

There are a variety of strains of \textit{Salmonella enterica} serovar Typhi, all of which have varying degrees of resistance to antibiotics. Initially, most \textit{S. typhi} strains were susceptible to first-line antimicrobial therapy; however, in 2011, there was a drastic increase in a new group of multidrug-resistant strains of \textit{S. typhi}, known as Haplotype 58, or H58.\textsuperscript{24}

H58 is a specific genetic variant of \textit{S. typhi}, which arose through genetic mutations.

Unlike other strains of \textit{S. typhi}, H58 strains display more resistance to current treatment methods, such as Fluoroquinolones, ampicillin, and trimethoprim-sulfamethoxazole, which have often been used to treat typhoid fever. They also display a wide range of genetic diversity, as well as increased fitness and transmission, which has allowed strains of H58 to be associated with several large-scale outbreaks in recent years. The typhoid outbreak within Pakistan can be attributed to Haplotype 58, and despite public health methods being in play, the drug resistance and fitness of strains of H58 have allowed it to rapidly spread and lead to such outbreaks.

\textbf{Molecular Structure and Drug Resistance of \textit{S. typhi} Haplotype 58}

Analyzing the molecular and genetic structure of typhoid is essential to understanding its effectiveness in spreading so quickly within Pakistan. As mentioned, the salmonella bacterium is transmitted directly through food and water. Salmonella is characterized as a flagellated facultatively anaerobic bacilli.\textsuperscript{25} Salmonella is able to reach the intestinal epithelium, surviving through the gastrointestinal tract. It can survive the acidic environment of the stomach, largely in part to its ability to
produce acid shock proteins and its strong outer membrane and cell wall. The outer membrane of Salmonella is asymmetric, with the inner layer consisting of a phospholipid layer and the outer layer a lipopolysaccharide layer, which is divided into three regions: lipid A, core, and O polysaccharides. This complex outer region of Salmonella has been suggested to have provided high virulence since the polysaccharide layer has a number of benefits, including but not limited to retaining water, providing protection against antibiotics, and containing efflux pumps that can remove antibiotics Jamilah. This gives Salmonella a high level of survivability, which has been essential in allowing it to spread throughout Pakistan.

Haplotype 58 has the above structural characteristics, but there are a number of genetic mutations that have provided it with even more increased resistance. These mutations consist of genes that code for polysaccharides and a more effective typhoid toxin, which causes typhoid fever. Furthermore, Haplotype 58 strains have undergone a mutation that causes increased expression in the rpoS gene, which is a stress response regulator. This allows increased survivability in extreme environments and more resistance to antibiotics due to the release of pertinent proteins for protection. To conclude, Haplotype 58 has expressed a variety of mutations, many not yet understood, that have directly led to increased transmission and drug resistance, which has made it difficult to contain its spread within Pakistan and provide potential treatments to the people.

Conclusions

Treatment, Preventative Measures, and Recommendations

Vaccination is at the forefront of efforts attempting to combat XDR-typhoid fever cases in Pakistan. Vaccines can be administered to children who are 6 months old with long-lasting effects, which is quite relevant, given that 60–70% of typhoid-related cases and deaths in Pakistan were in children under the age of 15, per a 2017 study. The efficacy of the vaccines is also undoubtedly evident, as the number of cases of typhoid fever in unvaccinated children was over double that in vaccinated children. Furthermore, according to the information obtained from Pakistan, the typhoid fever vaccine is 95% effective directed at Salmonella typhi strains that have been verified by culture, and it is 97% effective versus strains that have become resistant to the vaccine. Such efficacy rates underline the vaccine’s potency and may contribute to a decrease in the overuse of antibiotics that caused the establishment of several different strains of typhoid. XDR was initially noted to be within the southern province of Sindh, where a two-week vaccination program had been implemented in 2019 with the goal of immunizing 10 million children between the ages of 9 months and
15 years old. Two years later in Punjab, the subsequent stage of the vaccination drive was started with the goal of immunizing upward of 6.6 million kids who fell within a similar age range.\textsuperscript{30} The success of these efforts being championed by the WHO in tandem with the Pakistani government is undoubted, as multidrug-resistant typhoid cases have decreased by two-thirds-fold in Sindh and rates of vaccination in Punjab accelerated (World Health Organization, 2021).\textsuperscript{31}

With this critical information in mind, a couple of recommendations can be sorted (see Table 1).

The scope of these recommendations varies, highlighting the collaborative efforts required to tackle the public health crisis effectively. Basic hygiene practices are emphasized at the individual level, promoting personal responsibility in preventing transmission. At the local government level, collaboration with state governments and hospitals is suggested to disseminate awareness about typhoid and its preventive measures, indicating a more community-oriented approach. NGOs play a vital role in multiple areas, including distributing hygiene information and supplies, supporting healthcare workers in rural areas, and investing in telehealth systems for remote access to healthcare services. Moreover, local governments and NGOs share responsibilities in initiatives like constructing sanitation centers and distributing chlorine tablets, targeting improved sanitation and hygiene facilities in both urban and rural areas. These recommendations collectively underscore the need for a multifaceted approach, with contributions from various entities, to combat XDR-typhoid effectively and are the first steps toward a longer strategy aimed to combat and reduce the burden of disease spread found in Pakistan.\textsuperscript{12} Mass vaccination campaigns continue to be expensive, placed at $1.50 per dose, and need a stronger government and healthcare force at the federal and provincial levels for successful implementation. Social media posts also have many restrictive ends to them, given many of the impoverished (who are more susceptible to typhoid fever) do not have access to technology and applications in the way that others of a higher socioeconomic class are. Non-governmental organizations such as Coalition Against Typhoid (CaT), marked in financial transparency, continue to operate and promote awareness of typhoid fever’s spread, yet these efforts do not meet the minimum threshold to significantly reduce Pakistan’s prevailing public health crisis. A robust and forward-thinking approach needs to be taken by the Pakistani government to use its budget and international funding in the sectors of health and education to ensure disease prevention. It is crucial to understand that XDR-typhoid is a symptom of larger issues with global health security rather than a stand-alone issue. Antibiotic-resistant illnesses are an increasing global health challenge; thus, efforts to control XDR-typhoid must be viewed as a component of a bigger plan to address this issue. This entails funding the discovery and advancement of novel therapies, encouraging prudent antibiotic usage, and bolstering health
Table 1: Recommendations for XDR-Typhoid Prevention and Management

<table>
<thead>
<tr>
<th>Recommendations</th>
<th>Individuals</th>
<th>Local Government</th>
<th>NGOs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic hygiene practices</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Implementing health classes on typhoid fever in schools</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Collaboration with state governments and local hospitals to spread awareness</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Public postings on the effects of typhoid fever and preventative measures</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>The National Assembly should pass bills that raise the threshold for physicians employing prescriptions</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Azithromycin should be reserved for a case-by-case basis</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Allocating bonuses for healthcare workers in rural areas to incentivize working in these communities</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Investing in an integrated accessible telehealth system to allow reduction in costs of transport from remote areas</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Increasing the supply of diagnostic tools and testing kits</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Passing stronger laws on the prohibition of open defecation</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Government-subsidized workshops on constructing water filters, with priority in rural areas</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Long-term investment in the education sector to ensure meals and other resources are provided to children</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Construct sanitation centers in publicly accessible areas of Pakistan that are regularly maintained</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Distribution of chlorine tablets to slums and neighborhoods to ensure soap and cleaning products are available</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Recommendations</td>
<td>Individuals</td>
<td>Local Government</td>
<td>NGOs</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>Developing an integrated system of sewage disposal that enables rainwater to be stored and harvester for local drinking</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Financing efforts for healthcare workers to go to different neighborhoods to provide an in person training on health practices</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 1: (Continued)

systems to enhance the provision of care for infectious illnesses. In addition, it is imperative to make certain that the effort to respond to XDR-typhoid is equitable and just, considering the requirements of populations who are vulnerable and those who live in places with a lack of resources. This necessitates a focus on social justice and health equity, as well as an understanding that some communities bear a disproportionate amount of the burden of sickness. The core causes of health disparities, such as poverty and prejudice, which contribute to the development of contagious illnesses like XDR-typhoid, must also be addressed.

References


Evaluation of CDM and RBM Methods to Estimate Small Q-matrices

Raphael Jeong-Hin Chin

Introduction

Cognitive diagnosis models (CDMs) are psychometric models that assess one’s mastery of latent skills being tested. CDMs provide detailed feedback, including the probability of mastering a certain topic. Owing to CDMs’ effectiveness in determining strengths and weaknesses in the topics to be tested, researchers in the field are becoming more aware of CDMs and “assessment for learning rather than assessment of learning” (Ravand & Robitzsch, 2015).

Multiple formulations of CDMs have been proposed in psychometric literature such as deterministic inputs, noisy “and” gate (DINA) (de la Torre, 2009), generalized DINA (GDINA) (de la Torre, 2011), and log-linear cognitive diagnosis models (LCDM) (Henson et al., 2009). There are multiple packages to fit different CDMs, such as the cdmTools and CDM packages (Nájera et al., 2022; Robitzsch et al., 2022). These packages help researchers use CDMs to learn more about the examinees’ latent attributes based on their responses.

An important component of CDMs is the Q-matrix that informs the dependency structure between the J test items and K latent attributes (Li et al., 2022; Xu & Shang, 2018) because the Q-matrix can be effectively used to design intervention strategies. An example Q-matrix is shown in Table 1. ‘1’ in the matrix means that Skill K is required for mastery of Item J. Thus, Q-restricted latent class models have gained popularity in fields such as educational proficiency assessments, psychiatric diagnosis, and many more disciplines (Xu & Shang, 2018). A well-known usage of CDMs is to study the dependency between...

Contact: Raphael Jeong-Hin Chin <jeonghin@umich.edu>

https://doi.org/10.3998/umurj.5506
mathematical questions (items) and their latent skills for the topic of fractions as shown in Table 2. Let the six attributes tested in this topic be:

1. Find the lowest common denominator.
2. Add fractions.
4. Multiply fractions.
5. Divide fractions.
6. Convert mixed numbers to improper fraction.

The first item (mathematical question) in the test is $\frac{3}{4} + \frac{1}{2}$, where “find the lowest common denominator,” “add fractions,” and “convert mixed numbers to improper fractions” (skills 1, 2, and 6) are required for this question to be answered correctly. Thus, the rows of the Q-matrix corresponding to this item will contain the vector $(1, 1, 0, 0, 0, 1)$ as shown in Table 2.

The Q-matrix plays an important role in CDMs because it can be used to categorize test items and design future assessments (Li et al., 2022). However, not all assessments can be explicitly specified with a Q-matrix. Even if there is an explicitly specified Q-matrix, the Q-matrix may not be accurate due to the following reasons: (i) design error by the assessment provider; and (ii) one test item may be linked to multiple attributes, but not all attributes are found and identified. For example, error (i) is committed in the second row of Table 2 because skill 6 is

<table>
<thead>
<tr>
<th>Questions</th>
<th>Skill 1</th>
<th>Skill 2</th>
<th>Skill 3</th>
<th>Skill 4</th>
<th>Skill 5</th>
<th>Skill 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2\frac{3}{4} + 1\frac{1}{2}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$2\frac{3}{4} - 1\frac{1}{2}$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$2\frac{3}{4} - 1\frac{1}{4}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Q-matrix Corresponds to Three Math Questions and Six Latent Attributes
not required to correctly answer $2\frac{3}{4} - 1\frac{1}{2}$. Thus, it is important to be able to learn more about the Q-matrix from the responses in order to have a better understanding of the relationship between the test items and latent variables’ attributes.

Models

In this paper, the models of interest are the deterministic inputs, noisy “and” gate (DINA) model, the generalized-DINA (GDINA) model, and the restricted Boltzmann machines (RBMs). These three models are used in this paper to perform the following:

I. Test the accuracy of RBMs used by Li et al. (2022) on the data generated with a small number of latent attributes $K \in \{3, 4, 5\}$.
II. Compare the outputs from (i) with the results from Xu & Shang (2018).
III. Compare the results generated from the “CDM” package with the results from (i) and (ii) (Robitzsch et al., 2022).

Deterministic Inputs, Noisy “and” Gate (DINA) Model

The DINA model assumes a conjunctive relationship among attributes, where it is necessary to possess all the attributes indicated by the Q-matrix for a positive response (Xu & Shang, 2018). For each cell of the Q-matrix, $q_{jk}$ is 1 if the $k^{th}$ attribute is required to correctly answer the $j^{th}$ item. In this model, an examinee’s skills vector and the Q-matrix produce a latent response vector $\eta_i = \{\eta_{ij}\}$, where

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{jk}^{\eta_{ij}}$$

has a value of 1 if examinee $i$ possesses all the skills required for item $j$ or has a value of 0 if the examinee lacks at least one of the required skills (de la Torre, 2009). $K$ here represents the number of latent skills. Let $R_{ij} = \{0, 1\}$ represent the examinee $i$ answering item $j$ correctly. The uncertainties in this model are the slipping parameter, $s_j$, and guessing parameter, $g_j$, where

$$s_j = P(R_{ij} = 0 | \eta_{ij} = 1)$$

$$g_j = P(R_{ij} = 1 | \eta_{ij} = 0).$$
Therefore, the probability of examinee $i$ with skills vector $\alpha_i$ answering item $j$ correctly is given by

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = g_j^{1-n} (1-s_j)^{n}.$$  

**Generalized-DINA (GDINA) Model**

Similar to the DINA model, the GDINA model requires a $J \times K$ Q-matrix as well. For each cell of the Q-matrix, $q_{jk}$ is 1 if the $k^{th}$ attribute is required to correctly answer the $j^{th}$ item. In addition, GDINA separates the latent classes into $2^{K_j}$ latent groups where $K_j = \sum_{k=1}^{K} q_{jk}$ represents the number of required attributes for item $j$ (de la Torre, 2011). Let $\alpha_{ji}^*$ be the reduced attribute vector whose elements are the required attributes for item $j$, and then the probability that examinees with attribute pattern $\alpha_{ji}^*$ will answer item $j$ correctly is denoted by

$$P(X_j = 1 | \alpha_{ji}^*) = P(\alpha_{ji}^*).$$

In the GDINA model, there are three types of link functions available. This paper focuses only on the identity link function given by

$$P(\alpha_{ji}^*) = \beta_{j0} + \sum_{k=1}^{K_j} \beta_{jk} \alpha_{ik} + \sum_{k'=k+1}^{K_j} \sum_{k=1}^{K_j-1} \beta_{jk\k'} \alpha_{ik} \alpha_{ik'} + \beta_{j12...K_j} \prod_{k=1}^{K_j} \alpha_{ik}$$

where

- $\beta_{j0}$ is the intercept for item $j$;
- $\beta_{jk}$ is the main effect due to $\alpha_k$;
- $\beta_{jk\k'}$ is the interaction effect due to $\alpha_k$ and $\alpha_{k'}$; and
- $\beta_{j12...K_j}$ is the interaction effect due to $\alpha_1, ..., \alpha_{K_j}$.

2.3 **Restricted Boltzmann Machines (RBMs)**

RBMs are generative stochastic artificial neural network models that can learn probability distributions over a collection of inputs. RBMs were initially invented
by Paul Smolensky under the name Harmonium (Smolensky, 1986). RBMs used in this paper follow the model design in Li et al. Visible units are denoted by $R = \{R_1, \ldots, R_J\} \in \{0,1\}^J$, and hidden units are denoted by $\alpha = \{\alpha_1, \ldots, \alpha_J\} \in \{0,1\}^K$. RBMs are characterized by the energy functions with their joint probability distribution given by

$$P(R, \alpha; \theta) = \frac{1}{Z(\theta)} \exp\{-E(R, \alpha; \theta)\} \quad (2)$$

where $Z(\theta)$ is the partition function given by

$$Z(\theta) = \sum_{R \in \{0,1\}^J} \sum_{\alpha \in \{0,1\}^K} \exp\{-E(R, \alpha; \theta)\} \quad (3)$$

and $E(R, \alpha; \theta)$ is the energy function given by

$$E(R, \alpha; \theta) = -b^T R - c^T \alpha - R^T W \alpha = -\sum_{j=1}^J R_j b_j - \sum_{k=1}^K \alpha_k c_k - \sum_{j=1}^J \sum_{k=1}^K R_j w_{j,k} \alpha_k \quad (4)$$

In Equations 2–4, $\theta = \{b, c, W\}$ are the model parameters, $b \in \mathbb{R}^J$ are visible biases, $c \in \mathbb{R}^K$ are hidden biases, and $W \in \mathbb{R}^{J \times K}$ is the weight matrix describing the interactions between the visible and the hidden units. The hidden and visible units are conditionally independent as there are no “R-R” or “$\alpha - \alpha$” interactions (Li et al., 2022). $w_{j,k} \neq 0$ in the weight matrix, $W$, for RBMs indicates the presence of interaction between the visible and the hidden units. Although DINA and GDINA models violate the conditionally independent assumptions of RBM, it was shown in Li et al. that the Q-matrices for these models are estimable.

**Data**

The data are simulated with latent attributes dimension $K \in \{3, 4, 5\}$ and the number of test items, $J = 20$. The true Q-matrices chosen are identifiable and similar to those used in Xu & Shang (2018). The three true Q-matrices are
In this study, the data is simulated from the DINA latent class model. The ground truth response probabilities for all items are between 0.2 and 0.8, and both the slipping and the guessing parameters are set to 0.2. The dependency of latent attributes, $\rho$, is set to $\rho \in \{0,0.15,0.25,0.5\}$. The two-step simulation of true latent profiles follows those set in Xu & Shang (2018). First, $x_i$ is generated with $x_i=(x_{i1},...,x_{ik}) \sim N(0,\Sigma)$ for $i=1,...,N$ where $\Sigma=(1-\rho)I_k+\rho 1_k1_k^T$. The attribute profile $a_{ik}$ is set to be 1 if $x_{ik} \geq 0$ and 0 otherwise. The response data is then generated using the ‘sim.din’ function from the CDM package.

### Estimating the Q-matrix

The Q-matrices are estimated using the gdina function from the CDM package. As the response data follows the DINA model, a GDINA model can be fitted as the GDINA model is a generalized version of the DINA model. The gdina function will be used to fit the response data using both LASSO and Truncated
LASSO Penalty (TLP). The delta matrix returned by the function will be converted into a $J \times (2^K - 1)$ binary matrix (intercept column removed). The idea behind this is that because $\delta = \beta \times q$, if $\delta$ is not 0, $q$ is definitely not 0, where $\beta$ and $q$ are elements in Equation (1). Values that are close to 0 in the delta matrix (smaller than 0.1) will be forced to be 0 and everything else to be 1. The $J \times 2^K$ binary matrix will be collapsed into a $J \times K$ binary matrix by grouping up the latent attributes that are required to master the item $J$.

Let $\alpha \in \{0, 1\}$, $1 \leq k \leq K$, and $\delta_{ji} = \alpha_{jk}...\alpha_{jk}$ be the binary representation index of $i$\textsuperscript{th} element in the $j$\textsuperscript{th} row of the delta matrix. $\delta_{ji}$ will be transformed to have a value of 1 if it is greater than the threshold and 0 otherwise.

$$t_{jk} = \sum_{i=1}^{K} \delta_{ji} \text{ where } \alpha_{jk} = 1$$

(6)

$$\widehat{Q}_{jk} = 1 \text{ iff } t_{jk} \neq 0$$

(7)

For example, let $\delta = (1.4, 1.32, 0.08, 2.1, 0.0003, 0.0001, 0)$, $J = 1$, $K = 3$, and threshold = 0.1, then applying Equations (6), we get,

$$\delta = (1.5, 1.7, 0.01, 2.9, 0.008, 0.0021, 0) \Rightarrow (1, 1, 0, 1, 0, 1, 0)$$

$$t = (2, 2, 0)$$

In Equation (6), the columns of the $J \times (2^K - 1)$ binary matrix refer to (Attr1, Attr2, Attr3, Attr12, Attr13, Attr23, Attr123). The matrix is then collapsed into a $J \times K$ matrix by summing up all the 1s into their respective latent attributes, where the columns refer to (Attr1, Attr2, Attr3). The estimated Q-matrix in Equation (7) is expected to be identifiable only up to rearranging the orders of the columns. This is because when estimating the Q-matrix, the columns do not contain information about the latent attributes (e.g., the $n$\textsuperscript{th} column of the Q-matrix might not refer to the $n$\textsuperscript{th} latent attribute). Thus, the estimated Q-matrix will be reordered so that each column shows the lowest possible average congruent coefficient with the True Q-matrix’s columns. This process is done using the “orderQ” function in cdmTools (Nájera et al., 2022).

Accuracy Measurement

To evaluate the estimation accuracy, the entry-wise overall error (OE), out-of-true positive percentage error (OTP), and out-of-true negative percentage error (OTN) are reported. Their formulae are as follows:
\[
OE = \frac{1}{JK} \sum_{j=1}^{I} \sum_{k=1}^{K} \mathbb{1}\{\hat{q}_{j,k} \neq q_{j,k}\} \quad (8)
\]
\[
OTP = \frac{\sum_{j=1}^{I} \sum_{k=1}^{K} \mathbb{1}\{\hat{q}_{j,k} = 0, q_{j,k} = 1\}}{\sum_{j=1}^{I} \sum_{k=1}^{K} \mathbb{1}\{q_{j,k} = 1\}} \quad (9)
\]
\[
OTN = \frac{\sum_{j=1}^{I} \sum_{k=1}^{K} \mathbb{1}\{\hat{q}_{j,k} = 1, q_{j,k} = 0\}}{\sum_{j=1}^{I} \sum_{k=1}^{K} \mathbb{1}\{q_{j,k} = 0\}} \quad (10)
\]

**Results**

In this study, the Q-matrices are estimated completely from the data. For \( K = 3 \), the following crossover design is applied for the DINA model, three sample sizes, and four attribute-dependent levels: \( \{\text{DINA}\} \otimes \{N = 500, 1000, 2000\} \otimes \{\rho = 0, 0.15, 0.25, 0.5\} \). For \( K = 4 \) and \( K = 5 \), the designs are \( \{\text{DINA}\} \otimes \{N = 1000, 2000\} \otimes \{\rho = 0, 0.15, 0.25, 0.5\} \).

Table 3 shows the simulation results for 50 replications. From Table 3, it can be observed that, on average, TLP and RBM outperform the LASSO method. In the case of small \( N \) and small \( \rho \), RBM usually outperforms the TLP method. However, as \( N \) and \( \rho \) become larger, the TLP method is able to estimate a Q-matrix more similar to the true Q-matrix. As \( N \) increases, accuracy also increases. This is because if there is more response data, the model has more data to learn and train from, resulting in a more accurate model. Surprisingly, across the three

<table>
<thead>
<tr>
<th>( K )</th>
<th>( N )</th>
<th>Model</th>
<th>( \rho = 0 )</th>
<th>( \rho = 0.15 )</th>
<th>( \rho = 0.25 )</th>
<th>( \rho = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>500</td>
<td>Lasso</td>
<td>0.8253</td>
<td>0.8587</td>
<td>0.8610</td>
<td>0.8613</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.8420</td>
<td>0.8650</td>
<td>0.8823</td>
<td>0.9023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8420</td>
<td>0.8727</td>
<td>0.8957</td>
<td>0.9017</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>Lasso</td>
<td>0.9043</td>
<td>0.9117</td>
<td>0.9217</td>
<td>0.9220</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.9037</td>
<td>0.9453</td>
<td>0.9450</td>
<td>0.9593</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8667</td>
<td>0.9123</td>
<td>0.9323</td>
<td>0.9400</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Lasso</td>
<td>0.9300</td>
<td>0.9703</td>
<td>0.9667</td>
<td>0.9550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.9623</td>
<td>0.9813</td>
<td>0.9857</td>
<td>0.9930</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8893</td>
<td>0.9390</td>
<td>0.9440</td>
<td>0.9513</td>
</tr>
</tbody>
</table>

Table 3: Mean Accuracy (50 Repetitions) for \( K = 3, 4, 5 \), and \( J = 20 \)
Table 3: (Continued)

<table>
<thead>
<tr>
<th>K</th>
<th>N</th>
<th>Model</th>
<th>Accuracy (1-Error)</th>
<th>ρ = 0</th>
<th>ρ = 0.15</th>
<th>ρ = 0.25</th>
<th>ρ = 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1000</td>
<td>Lasso</td>
<td>0.7375</td>
<td>0.7930</td>
<td>0.8030</td>
<td>0.8220</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.7528</td>
<td>0.8140</td>
<td>0.8515</td>
<td>0.8740</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8285</td>
<td>0.8395</td>
<td>0.8588</td>
<td>0.8970</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Lasso</td>
<td>0.8323</td>
<td>0.8615</td>
<td>0.8738</td>
<td>0.8673</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.8453</td>
<td>0.8918</td>
<td>0.9008</td>
<td>0.9185</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8553</td>
<td>0.8708</td>
<td>0.8928</td>
<td>0.9093</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>Lasso</td>
<td>0.6500</td>
<td>0.6648</td>
<td>0.7006</td>
<td>0.7452</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.6500</td>
<td>0.6784</td>
<td>0.7188</td>
<td>0.7736</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8282</td>
<td>0.8534</td>
<td>0.8404</td>
<td>0.8432</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Lasso</td>
<td>0.7096</td>
<td>0.7768</td>
<td>0.8122</td>
<td>0.8348</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TLP</td>
<td>0.7228</td>
<td>0.8004</td>
<td>0.8330</td>
<td>0.8990</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RBM</td>
<td>0.8668</td>
<td>0.8730</td>
<td>0.8850</td>
<td>0.8714</td>
<td></td>
</tr>
</tbody>
</table>

different methods, the accuracy increases as the correlation among attributes increases. This may be because the higher the dependency among the attributes, the lesser the number of possible attribute patterns, making estimation relatively easier (Li et al., 2022).

Conclusion and Future Direction

In conclusion, it is shown in Table 3 that the CDMs with TLP method outperformed the ones with LASSO method. Moreover, it is interesting to see that the RBM models have stable performance for \( K \leq 5 \). The RBM models always have an accuracy of 82% or more for these data while CDMs perform badly when \( N \) is small.

The future work of interest would be to explore different ways to include interactions between latent attributes so that the assumptions set in RBM will not be violated. In practice, it is hard to find latent attributes that do not correlate with one another. Thus, by addressing this latent attribute interaction problem, the RBM method that has higher accuracy can be created. One potential way to address this problem may be integrating deep learning into the RBM method.

Owing to the success and stability of the RBM method in learning dichotomous item responses, it will also be interesting to implement the RBM method in
research that uses polytomous item responses. This is because a lot of questionnaires contain responses in the form of a 5-point or 7-point Likert scale. It will be interesting to study how different levels of responses correlate with mastering a certain skill or how the slipping and guessing parameters are affected by the way the questions were phrased. For example, an examinee may have the skills to answer a mathematical question correctly, but because the questions contain ambiguity and poor word choices, the examinee may be unable to answer the question.

Acknowledgment

The author would like to thank Prof. Gongjun Xu and Dr. Chenchen Ma for their advice, guidance, and the opportunity to engage in this research, the Advanced Research Computing at the University of Michigan for their consultation on code performance, and Weihan Xu for his assistance.

References

Multimodal Imaging Evaluation of Rabbit Models

ASHLEY BROWN, VAN PHUC NGUYEN, IYABODE AJAYI, AND YANNIS M. PAULUS

Currently available imaging modalities each have limitations in their scope, preventing a comprehensive view of the eye. A multimodal system can simultaneously use numerous modalities to increase the effectiveness of imaging to provide structural and functional information while decreasing the time and invasiveness of multiple procedures. In this study, optical coherence tomography (OCT), photoacoustic microscopy (PAM), fundus photography, fluorescein angiography (FA), and indocyanine green angiography (ICGA) were performed on a baseline rabbit model to verify the imaging system. Clear images distinguishing retinal vessels, choroidal vessels, and retinal layers were obtained, verifying the multimodal system’s efficacy before imaging experimental models. Imaging macular degeneration and glaucoma animal models can advance understanding of novel treatments in ophthalmic research. Multimodal imaging also offers a promising new means for the early detection of retinal diseases in clinical settings.

Keywords
multimodal imaging, eye imaging, optical imaging, photoacoustic microscopy, optical coherence tomography, animal models

Introduction

Age-related macular degeneration (AMD) is the progressive aging of the macula, the part of the retina that allows you to see objects in your central vision. It is a leading cause of blindness and vision impairment in patients over 60 years old. As of 2019,
19.8 million people above the age of 50 in the United States have been diagnosed with AMD\(^1\). An additional study measured the direct cost of AMD diagnosis, treatment, and management in the United States at $10 billion per year\(^7\). This cost does not include indirect medical expenses, such as home health care, nursing homes, additional vision aids, or productivity losses from being unable to work. Vision loss from AMD is currently irreversible, as retinal photoreceptors lack a regenerative capacity. However, regenerative cell therapies can offer a promising future treatment in restoring retinal function through transplanted stem cells\(^1\).

Glaucoma is the second leading cause of blindness in the United States. As of 2020, an estimated 3 million Americans were living with glaucoma. Dangerously, around 50% of the 3 million people with this disease are undiagnosed due to little to no symptoms at the disease onset\(^3\). Symptoms of permanent vision loss tend to occur during late-stage disease progression once the patients have irreversible tunnel vision. In glaucoma, the aqueous humor, an optically clear fluid that nourishes and maintains the intraocular pressure (IOP) of the eye, begins to accumulate. This accumulation raises the IOP and irreversibly damages the optic nerve, causing progressive loss of peripheral to central vision. Current treatments involve topical eye drops or surgery to lower the IOP and prevent further vision loss. Unfortunately, these treatment options are not always financially accessible or fully effective. Topical medications can cost up to $874 per year depending on the medication brand and prescribed regimen\(^19\), while surgery costs are dependent on an individual’s insurance and come with increased risks of vision loss, bleeding, and infection\(^5\). Optical imaging could allow early disease detection and proactive treatment to slow or avoid vision loss completely.

Optical imaging is an important tool in clinical and research medical settings to diagnose patients and monitor treatment outcomes. Imaging can be used in the early-stage diagnosis and progressive monitoring of many eye diseases. In addition, longitudinal imaging can provide qualitative and quantitative insight on disease progression and evaluation of novel treatment techniques in developed animal models. There are currently a variety of available imaging systems; however, none can provide a high-resolution, three-dimensional image of the entire eye with a full anatomic and functional assessment. The most common techniques include fundus photography, fluorescein angiography (FA), indocyanine green angiography (ICGA), optical coherence tomography (OCT), and photoacoustic microscopy (PAM).

Fundus photography uses a low-power light microscope attached to a camera to provide a noninvasive, wide-field view of the interior surface of the eye\(^13\); however, it can be limited in image depth and cannot visualize the choroid. Fundus photography can be paired with FA to visualize the blood flow in the retina. This technique is commonly used to diagnose and monitor nonperfusion and
neovascularization\textsuperscript{12}. FA requires the intravenous injection of a fluorescent dye fluorescein sodium and thus is more invasive and holds risks of adverse side effects, including the risk of nausea, vomiting, anaphylaxis, and death.

ICGA allows visualization of the choroidal vessels using the exogenous contrast dye indocyanine green (ICG)\textsuperscript{11}. ICG is water-soluble and nontoxic. New blood vessels detected in the choroid could indicate macular degeneration or central serous chorioretinopathy\textsuperscript{13}. Similar to FA, ICGA and ICG require the injection of an exogenous dye and only provide two-dimensional information.

OCT uses light waves to noninvasively produce a high-resolution, cross-sectional image of the retinal layers. This technique can be used to measure depth and evaluate changes in morphology to diagnose retinal diseases\textsuperscript{12}. OCT has limited ability to penetrate and evaluate deeper structures, such as the choroid, and cannot visualize leakage like FA and ICG can. OCT angiography (OCTA) not only can visualize microvascular structures in the retina but also has limited ability to evaluate the choroidal vessels\textsuperscript{24}. Both OCT and OCTA currently have limited fields of view, although imaging technologies are expanding rapidly.

PAM is a nonionizing, noninvasive technique that utilizes a short-pulsed laser to produce light\textsuperscript{12}. The absorption of this light generates a sound that can be detected to create a high-resolution, high-depth image of deep biological tissues. Each imaging technique has a variety of beneficial applications, but none can individually produce a high-resolution, three-dimensional image of the eye with a full anatomic and functional assessment. A multimodal imaging system could combine numerous imaging techniques, offsetting their individual limitations, to generate a comprehensive image of the eye. This would allow advanced early detection and progressive disease tracking in clinical and research settings.

**Methods**

To evaluate the rabbit model, this study used a multimodal imaging system consisting of OCT, PAM, fundus, FA, and ICGA. The use of a rabbit model was selected over the use of other rodent models as rabbits are phylogenetically closer to humans. Rabbit and human eyes share similarities in size, internal structures, choroidal blood supply, vitreous volume, and biochemical features. In addition, rabbit eyes have been previously studied to verify standardized values of fluid flow across the retina and retinal pigment epithelium (RPE). This paper focuses on imaging a baseline model, a group known to have no mutations or diseases, to obtain control data. The study was continued on both AMD and glaucoma experimental models following baseline system verification.
Animal Model

One New Zealand white rabbit (female, 9 months of age, weight 4 kg) was obtained for this study. All experiments were performed in compliance with the guidelines set by the Association for Research in Vision and Ophthalmology (ARVO) after the approval of protocol PRO00010288 from the University of Michigan’s Institutional Animal Care and Use Committee (IACUC).

The rabbit was anesthetized 15 minutes prior to imaging with an intramuscular injection of ketamine (40 mg/kg) and xylazine (5 mg/kg). A local anesthetic of 0.5% topical proparacaine eye drops was also administered. Ten minutes prior to imaging, rabbit eyes were dilated with tropicamide 1% ophthalmic and phenylephrine hydrochloride 2.5% ophthalmic solution. The rabbit was kept on a water-circulating heat blanket to maintain body temperature. Vitals, including mucus membrane color, heart rate, respiratory rate, anal temperature, and ambulation, were monitored every 15 minutes through induction and maintenance of anesthesia until the rabbit was fully recovered and alert.

AMD and Glaucoma Model Creation

After baseline imaging, AMD and glaucoma were induced in New Zealand white rabbits through a photocoagulation laser pulsed at 50 ms with a 300 μm diameter that can be used to create localized lesions to the retinal pigment epithelium (RPE) to represent cell atrophy causing vision loss similar to AMD. In the glaucoma model, 4 mg of triencede steroid (40 mg/ml) was injected intravitreally into the vitreous to increase the IOP through a steroid response mechanism.

OCT and PAM

Baseline imaging was conducted and then repeated longitudinally for 30 days following model creation. Anesthetized rabbits were positioned so that the imaging system could observe the area of interest. OCT images were obtained at an acquisition rate of 36 kHz with a lateral resolution of 3.8 μm, axial resolution of 4 μm, and a depth of 1.9 mm using the Ganymede-II-HR OCT system (Thorlabs, Newton, NJ)\textsuperscript{11}. To obtain PAM images, a custom-built ultrasonic transducer was placed on the conjunctiva of the eye. Saline was administered to decrease corneal dehydration and to allow ultrasound signal coupling. PAM images were acquired at 578 nm with 80 nJ, half of the ANSI safety limit of 160 nJ
at 578–650 nm using an OPO nanosecond pulsed laser light (3–5 ns) with a pulse repetition rate of 1 kHz pumped by diode solid-state laser (NT-242; Ekspla, Vilnius, Lithuania). PAM images had a lateral resolution of 4.1 μm and an axial resolution of 37 μm.12

**Fundus, FA, and ICGA**

Baseline imaging was conducted and then repeated longitudinally following model creation. Fundus photography was conducted with the Topcon 50EX imaging system. About 0.2 ml of 10% fluorescein sodium was injected intravenously in the marginal ear vein for FA. Images were taken immediately following the injection. Saline was topically administered to decrease corneal dehydration. For ICGA, 100 μl of ICG contrast (25 mg/3 ml) was delivered intravenously in the marginal ear vein immediately prior to imaging.

**Image Analysis**

All images were analyzed with Image J and then compiled into a three-dimensional model using Amira software.

**Results**

Figure 1D depicts the retinal veins through fundus photography. The optic nerve can be visualized on the left side of the image. The fundus image provides insight into the morphology of the retinal vein, running from the bottom left to the top right of the image; however, we are unable to analyze the choroidal vessels. The subsequent FA and ICGA images are better able to visualize the blood flow in the choroid, as seen in the dense region of smaller vessels running perpendicularly to the large retinal vein. In Figure 1B, FA demonstrates the retinal veins, along with the surrounding choroidal vessels. Similar to FA, ICG was also used as an injected contrast agent to visualize the retinal and choroidal vessels on a 2D plane in Figure 1C.

Figures 1E and 1F present a 2D cross section of the retina using OCT. Similar to fundus photography, the OCT images successfully visualize the retinal layers but are limited in their view of the choroidal vessels. The structural layers of the retina, including the RPE, choroid, and sclera, are clearly visualized. At 2.6, 3.1, and 3.4 mm on the x-axis of image x, the retinal veins depicted in Figures 1A–D are visible. Figure 1A shows the x, z plane of these
two images compiled into a 3D OCT scan. The scan spatially visualizes the retinal veins with respect to the choroidal veins.

Discussion

As mentioned in the introduction, each individual modality has its own limitations. Alone, fundus photography is restrained by a lack of image depth and inadequate visualization of the choroidal vessels. When fundus is paired with FA, we can better detect retinal blood flow. Figure 1B shows increased detail in the retinal vein compared to only fundus in Figure 1D. Similarly, when pairing fundus with ICGA, we now observe choroidal blood flow. The brighter-colored area behind the retinal vein in Figure 1C indicates healthy choroidal vessels. All of these modalities lack visualization of retinal layers, so OCT was needed to
observe in vivo morphology. The digital combination of the images from each modality forms the 3D digital construction of the retinal vessel and choroidal vessels seen in Figure 1A. The combination of these imaging systems offsets their individual limitations to form a comprehensive image that no individual system can currently capture.

Longitudinal imaging with this multimodal system can be conducted to monitor for retinal disorders, such as AMD and glaucoma. The disruption of blood flow from retinal and choroidal vein occlusions (RVO and CVO) can lead to the creation of new blood vessels (neovascularization) or vessel leakage. Both neovascularization and leakage are indicators of retinal disorder. For example, an increase in neovascularization of the choroid can often indicate progression of AMD. Early disease diagnosis allows preventative treatment before the onset of symptoms, including irreversible vision loss.

**Conclusion**

The combination of OCT, fundus photography, FA, and ICGA in the reported images effectively provides a comprehensive image of the retinal model. This multimodal technique offsets the limitations of individual modalities and combines their strengths in an efficient system. The increased detail of these overlaid images could assist clinicians in early retinal disease detection, prior to symptom onset, quickly and noninvasively. In addition, a verified disease model would allow the future exploration of novel treatments, such as regenerative therapies.

**Funding Information**

This work was supported by the Undergraduate Research Opportunity Program (UROP) and was funded by grants from the National Eye Institute (YMP:1K08EY027458, 1R01EY033000, 1R41EY031219), Fight for Sight—International Retinal Research Foundation (YMP: FFSGIA16002), Alcon Research Institute Young Investigator Grant (YMP), unrestricted departmental support from Research to Prevent Blindness, generous support of from Helmut F. Stern Career Development Professorship in Ophthalmology and Visual Sciences (YMP), and the University of Michigan Department of Ophthalmology and Visual Sciences. This research utilized the Core Center for Vision Research funded by the National Eye Institute (P30 EY007003). We thank Drs. Yuqing Chen and Dongshan Yang and the University of Michigan CAMTraST for their generous donation of rabbits.
References


An Analysis of SPEEK as a Suitable Alternative for Nafion in a Proton Electrolyte Membrane in PEM Fuel Cells

Ihsaan Yasin and Bilal Irfan

The Proton-Exchange Membrane Fuel Cell (PEMFC) has the potential to enable hydrogen as a clean energy storage medium worldwide. However, before it can be implemented globally, the PEMFC needs to be optimized with innovative technology to maximize its efficiency. This investigation analyzes the methods in which the efficiency can be optimized through a different polymer material that can be used as a solid polymer electrolyte membrane in the PEMFC, comparing Nafion with Sulfonated Poly (ether ether ketone) as the most effective electrolyte membrane by efficiency.

Introduction

The demand for efficient, renewable, and modern-day energy is more important than ever. Individually, energy is a fundamental part of today’s society, empowering the majority of our infrastructure. Currently, the majority of energy is produced by burning fossil fuels, comprising 80% of the world’s energy consumption in 2019 (Total energy consumption, 2021). And coupled with the declining quantity of fossil fuels, readily available methods of creating efficient, renewable energy are becoming a more important concern to be addressed. One potential technology that can aid in solving this issue is the Hydrogen Fuel Cell (HFC), a device that converts hydrogen fuel’s chemical energy into electrical energy. (Brandon and Thompsett, 2005: 59). Specifically, the Proton-Exchange Membrane Fuel Cell (PEMFC), which is more compact than other HFCs, has the potential to be implemented globally as a clean power source. Before its implementation, the PEMFC needs to have maximal efficiency to ensure easy integration into our daily use.

Contact: Ihsaan Yasin <iyasin@umich.edu>

https://doi.org/10.3998/umurj.5508
This investigation analyzes the methods in which the PEMFC’s efficiency can be optimized through a solid polymer electrolyte material in the PEM, considering the current most used polymer’s PEMFCs.

Background

The PEMFC has the means of creating clean and reliable energy. It is also very compact and quiet during its operation. Furthermore, the by-products produced as a result of the chemical reactions are environmentally friendly and can have a crucial effect in mitigating the rate of global warming (Brandon and Thompsett, 2005: 377).

The main component within a PEMFC that creates energy is called the Membrane Electrode Assembly (MEA). It is composed of four key parts: solid polymer electrolyte membrane, the electrodes, the catalyst layers, and the gas diffusion layer. The electrodes consist of an anode and a cathode separated by the PEM. At the anode, hydrogen is chemically reduced, producing protons, electrons, and heat (Thompson). The reaction is accelerated with the aid of a platinum catalyst, given that the operation temperatures of a PEMFC do not allow for electrochemical reactions to occur at a rapid rate (Zhou and Chen, 2012).

\[
\text{Anode Reaction: } H_2 \rightarrow 2H^+ + 2e^- \\
\]

Once the hydrogen is reduced, the protons migrate across the PEM, while the electrons enter the electrical circuit through the anode. The anode absorbs the electrons, which provides it with a negative charge. The electrons will travel from the negatively charged anode to the positively charged cathode. This migration produces voltage difference, which produces electricity. Concurrently, protons are traversing the PEM to the other side. On the cathode side, oxygen is present, hydrogen comes in through the membrane, and electrons exit the electrical circuit through the cathode. With these particles, the oxygen will be oxidized at the cathode to produce water.

\[
\text{Cathode Reaction: } O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \\
\]

Proton-Exchange Membrane

The PEM is an important component of the fuel cell. As previously mentioned, it allows protons to permeate to the cathode side. This characteristic is known as proton conductivity (Crofts, Antony, 1996). The PEM is unique compared to
other fuel cell membranes, since it is a solid polymer electrolyte. This contrasts other fuel cells, which have liquid-based electrolytes composed of a fluorinated sulfonic acid polymer called Nafion (Brandon and Thompsett, 2005: 61). A benefit of using Nafion is that it can have excellent proton conductivity. In addition, the cathode on the other side of the membrane further aids in attracting the protons, increasing the PEMFC efficiency.

**Nafion**

Currently, one of the most common types of polymers for the electrolyte membrane is Nafion, a perfluorinated sulfonic acid ionomer (Rieke et al, 2001). As a polymer, Nafion is inhomogeneous, displaying hydrophobicity and hydrophilicity. Proton conduction is carried out by the hydrophilic, negatively charged sulfonate group, which contains one hydrogen, as seen in the next figure (Rieke et al., 2001). The group’s negative charge attracts protons, and, with the Grotthuss mechanism, they can jump across the sulfonic acid group donors until the protons reach the negatively charged cathode (Sun et al., 2015). With this
characteristic, Nafion has proved to be effective, providing efficient proton conductivity and membrane lifetimes of over 60,000 hours (Hui San Thiam et al). In order to maintain this efficiency, Nafion has two important requirements.

First, it is only able to conduct protons when hydrated. In low-humidity environments, its effectiveness rapidly declines despite the aforementioned Grotthuss mechanism that allows proton-jumping in hydrogen. This is attributed to the fact that under anhydrous conditions, it is more difficult for the essential hydrogen bonds to form on the sulfonic acid groups, hence making it less permeable (Brandon and Thompsett, 2005). Water lowers Nafion’s tightness, which helps its membrane structure to adapt to allow faster proton conductivity. Furthermore, water in itself is already effective as a proton conductor. Water molecules are also held together by hydrogen bonds, creating a network of hydrogen bonds that can be used by protons across the network, as seen in the next figure (Sun et al., 2015). In terms of the Nafion membrane, water molecules work in conjunction with the sulfonic acid groups. They create bridges between each of the sulfonic acid groups, allowing for protons to easily diffuse across the Nafion membrane. Thus, maintaining a consistent humidity within the PEMFC is extremely important, as having inadequate humidity will drastically impede proton conductivity.
Another important factor of a Nafion-based membrane is that temperature can also have an important effect on the membrane’s ability to conduct protons. In general, higher temperatures are desirable in the PEMFC since it allows for faster-moving particles. Ultimately, this leads to faster electrochemical reactions at both the anodes and the cathodes as well as faster proton conductivity (Zhou and Chen, 2012). However, a very high temperature is not achievable due to the limitations of the Nafion polymer. Nafion is dependent on liquid water for efficient proton conduction, which becomes an issue when higher temperatures decrease the humidity of the membrane. Thus, the Nafion membrane is limited to temperatures below 100°C (Zhou and Chen, 2012).

Overall, Nafion-based proton electrolyte membranes are limited by humidity and temperature. Manufacturers must consider this, meaning it becomes a difficult task to make the PEMFC more efficient. With other polymers that are not greatly limited by temperature and humidity, the PEMFC could become drastically more efficient. This investigation focuses primarily on temperature and its effects on the polymer.

Due to the aforementioned limitations of Nafion, a more effective replacement for the Nafion-based membrane is required. There is one type of polymer that can potentially replace Nafion-based proton electrolyte membranes: sulfonated aromatic polymers.

Sulfonated Aromatic Polymers

Sulfonated aromatic polymers are one important type of material that could effectively be used as a replacement for Nafion-based fuel cell membranes. These polymers consist of an aromatic polymer backbone and a proton-conducting polymer electrolyte. Aromatic polymers by themselves are composed primarily of benzene rings or aromatic heterocyclic rings (Liu et al, 2012). The linkages between the rings can be both regular and rigid or irregular, making them branched and more flexible. This characteristic of aromatic polymers allows them to have desirable thermal properties in a fuel cell (Liu et al., 2012). This same acidic property is able to be applied to sulfonic acids. Aromatic polymers can be sulfonated through the combination with a concentrated solution of acids, such as sulfuric acid, chlorosulfonic acid, pure or complexed sulfur trioxide, and acetyl sulfate (Brandon and Thompsett, 2005). Thus, an aromatic polymer that has great thermal properties now gains a sulfonic acid group, creating a sulfonated aromatic polymer. This sulfonated aromatic polymer, then, has the ability to conduct protons. This is important for proton electrolyte membranes, and the ability of sulfonated aromatic polymers to conduct protons opens the potential for them to replace Nafion. However, a further investigation and analysis needs
to be done regarding their replaceability of Nafion-based membranes. One main polymer that could potentially replace Nafion is sulfonated poly (ether ether ketone). As a sulfonated aromatic polymer, it has a very strong backbone with benzene rings along with a sulfonic acid group that allows it to let protons pass through it with the Grotthuss mechanism.

**Thermal Stability**

Thermal stability of polymers is an extremely important part in creating an efficient proton electrolyte membrane; specifically, the ability to maintain structural integrity at high operating temperatures. Higher temperatures allow for a myriad of benefits: faster proton conductivity, low rates of catalyst poisoning, and little to no electrode flooding due to having water evaporate (Brandon and Thompsett, 2005: 382). Together, these properties prolong the life of a hydrogen fuel cell and give it more efficiency. This outcome is both cost-effective and efficient for the fuel cell.

Thermal stability can be measured in many ways. One main way of measurement is through Thermogravimetric Analysis (TGA). TGA is used to measure the weight of molecules as a function of time or temperature. It can be used to measure a compound’s weight as it heats or cools; however, for Nafion and Sulfonated Poly(Ether Ether) Ketone (SPEEK), the temperature will be increasing as a strong thermal stability at high temperatures is needed from polymer electrolyte membranes in fuel cells. The focus of measuring temperature is to understand how it affects the weight loss in polymers. High temperatures make polymers increasingly unstable, leading to decomposition. Despite high temperatures being good, bad thermal stability signifies a lower lifespan for electrolyte membranes, which is a major disadvantage for an effective electrolyte membrane.

The TGA data will be translated into Derivative Thermogravimetric (DTG), which is a graph that shows how the TGA values are changing at different
temperatures. DTG graphs are ones that represent the derivative of the TGA graphs. The peaks in DTG graphs let us understand the phase shifts of the polymer.

**Weight Percentage Data**

<table>
<thead>
<tr>
<th>Temperature 5°C</th>
<th>Percent Weight(%) 1</th>
<th>Temperature 1°C</th>
<th>Percent Weight 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>99.9</td>
<td>0</td>
<td>99.7</td>
</tr>
<tr>
<td>50</td>
<td>98.1</td>
<td>50</td>
<td>99.4</td>
</tr>
<tr>
<td>100</td>
<td>97.4</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>150</td>
<td>96.9</td>
<td>150</td>
<td>99.4</td>
</tr>
<tr>
<td>200</td>
<td>96.3</td>
<td>200</td>
<td>96.5</td>
</tr>
<tr>
<td>250</td>
<td>95.1</td>
<td>250</td>
<td>92.5</td>
</tr>
<tr>
<td>300</td>
<td>93.8</td>
<td>300</td>
<td>83.4</td>
</tr>
<tr>
<td>350</td>
<td>83.2</td>
<td>350</td>
<td>81.6</td>
</tr>
<tr>
<td>400</td>
<td>70.7</td>
<td>400</td>
<td>80</td>
</tr>
<tr>
<td>450</td>
<td>30.5</td>
<td>450</td>
<td>76.3</td>
</tr>
<tr>
<td>500</td>
<td>9.87</td>
<td>500</td>
<td>68.7</td>
</tr>
<tr>
<td>550</td>
<td>5.49</td>
<td>550</td>
<td>59.8</td>
</tr>
<tr>
<td>600</td>
<td>3.93</td>
<td>600</td>
<td>54.9</td>
</tr>
<tr>
<td>650</td>
<td>3.29</td>
<td>650</td>
<td>52.9</td>
</tr>
<tr>
<td>700</td>
<td>2.6</td>
<td>700</td>
<td>51.8</td>
</tr>
</tbody>
</table>

The Dagra Blue Leaf software will be used to extract raw values present in the graph. The data for both Nafion and SPEEK will be analyzed in intervals of 50°C, which provides a cohesive graph of how they fluctuate throughout the environments with increasing temperatures. The uncertainties from the literature data will be utilized.

After this extraction of TGA data, the derivative of the curve of best fit will be created. This allows a more in-depth analysis of where the different weight loss shifts occur within Nafion and SPEEK at high temperatures.

Nafion data has been taken from the source publication study, “The quintuple-shape memory effect in electrospun nanofiber membranes” (Zhang et al., 2013).

SPEEK data has been taken from the source publication study, “Electrospun multifunctional sulfonated carbon nanofibers for design…” (Li et al., 2017).
Nafion TGA

\[ y = (-1.14 \times 10^{-11})x^5 + (2.51 \times 10^{-8})x^4 + (1.81 \times 10^{-5})x^3 + (4.70 \times 10^{-3})x^2 - 0.401x + 103 \]

Nafion DTG
SPEEK TGA

\[ y = \left(3.71 \times 10^{-12}\right)x^5 - \left(6.19 \times 10^{-9}\right)x^4 + \left(3.81 \times 10^{-8}\right)x^3 \\
- \left(1.15 \times 10^{-3}\right)x^2 + 0.1x + 98.7 \]

SPEEK DTG

![Graph of Weight vs. Temperature of SPEEK](image)

![Graph of Rate of Change of Weight vs. Temperature of SPEEK](image)
Quantitative Interpretation for Thermal Data

According to Nafion’s TGA graph, Nafion’s final weight at 700°C is 2.6% of its initial mass. Further, the DTG graph points at which it equals 0, representing the peaks of Nafion. The two low peaks of Nafion are at temperatures of 63.3°C and 630°C, while the single high peak is at a temperature of 206°C. Nafion’s derivative TGA indicates a one high peak at 126°C, with a rate of weight change of 0.105% and one low peak at 426°C with a rate of weight change of −0.4%.

Shifting to SPEEK’s final weight percentage at 700°C is 51.8%. Its corresponding DTG graph showcases that the TGA graph of SPEEK contains a maximum peak at 58.7°C and a minimum peak at 676°C. Furthermore, the peak of the DGA graph is at a temperature of 529°C with a rate of weight change of −0.125%.

Thermal Data Analysis

The DGA graph provides a range of important information to understand the TGA graph and the thermal stabilities of Nafion and SPEEK. Initially, it can be seen from the TGA data that Nafion’s final weight percentage is 2.6% and SPEEK’s is 51.8%. However, this does not provide information about the usefulness of a polymer’s thermal properties in a fuel cell membrane, as even high temperatures do not operate at temperatures of 700°C. Thus, a more in-depth understanding can be gained by paying more attention to the DGA data. The peaks of the polymers in terms of weight loss were calculated using the derivative TGA graphs. We observe that Nafion exhibits two close peaks at 63°C and 206°C. These peaks are very similar and have a very small range of weight percentage of 92–100%; they do not indicate any major molecular phase shifts. SPEEK has an initial maximum of 58.663, which from the initial temperature has ranges of 98–100%. Though these initial peaks don’t showcase phase shifts, they showcase that Nafion and SPEEK are thermodynamically very similar at the initial temperature increases.

The following peaks calculated by the derivative graph occurs at 630°C, while SPEEK exhibits its minimum peak at 675°C. SPEEK’s last minimum peak occurs at 45°C higher than Nafion’s, indicating two things. SPEEK is potentially more stable as its peak occurs later than Nafion’s, showcasing that more temperature is required and, thus, more energy is required to break down bonds within SPEEK.

However, a stronger conclusion can be built with the peaks of the DGA graphs themselves. The peaks of the DGA graphs indicate inflection points on the TGA graphs. These inflection points will have the greatest rate of change
exhibited by the polymer. A great rate of change is a primary example of a phase shift in a polymer at a certain temperature. Phase shifts occur when a certain molecular part of the polymer starts to rapidly decompose under a specific temperature, and thus, the inflection points locate the same. The Nafion DGA graph indicates a main peak of \(-0.396\) at a temperature of \(426^\circ C\). Therefore, Nafion has a large phase shift in molecular decomposition with a very fast rate of decomposition. SPEEK indicated a phase shift at \(529^\circ C\), which occurred much later and had a much lower rate than Nafion at \(-0.125\). SPEEK reaches the greatest weight loss\% at \(100^\circ C\) later than Nafion, and it also has a much lower overall value. Molecularly, this can be attributed to the fact that SPEEK is more regular than Nafion and thus forms a tighter regular polymolecular structure. While the data showcases that Nafion exhibits its first weight loss phase at \(206^\circ C\) based on the TGA peak compared to that of SPEEK earlier at \(56^\circ C\), SPEEK demonstrates a much lower weight loss rate overall and enters its shift in phase much later than Nafion. Therefore, the aromatic backbone gives SPEEK more thermal stability.

The final thing that needs to be taken into account is the temperature range in which the sulfonic acid groups decompose in both Nafion and SPEEK. Despite established data that SPEEK is more thermally stable from \(0^\circ\) to \(700^\circ C\), sulfonic acid groups primarily provide both Nafion and SPEEK with their proton-conducting properties; their loss directly impacts fuel cell efficiency. Thus, they need to be taken into the greatest account. The initial DGA peaks showcased inflection points, which showcased the moment in which both Nafion and SPEEK experienced phases shifts in molecular decomposition. The decomposition occurring prior to this phase shift would be that of the sulfonic acid groups. Since these groups have weaker linkages compared to those of the backbones in Nafion and SPEEK, they decompose quicker. It may appear that the sulfonic groups are decreasing from \(0^\circ C\) to \(206^\circ C\) in Nafion and from \(0^\circ C\) to \(58^\circ C\) in SPEEK, but the rate of change of both these areas is very low. The rate thus indicates that the decomposition must be the water molecules that have been absorbed. Nafion has hydrophilic areas, and SPEEK is nearly fully hydrophilic, making them eligible to absorb water. Thus, this low decomposition can be fully attributed to water loss. After that, another drop in the rate of change occurs. As stated, the inflection point of Nafion is at \(426^\circ C\), and SPEEK’s is at \(529^\circ C\). The inflection point is based on a phase shift, and, due to the chemical changes that occurred, SPEEK’s decomposition must occur in the first phase of weight loss. This is due to the sulfonic acid group’s more unstable nature, causing it to decompose at earlier temperatures. The most crucial concept is that Nafion is moving into the second phase much earlier, thereby losing most of its sulfonic acid groups earlier than SPEEK while also at a faster rate of \(-0.4\) compared to \(0.125\) for SPEEK. This overall makes Nafion more thermodynamically stable than SPEEK.
Nafion could still be utilized at lower temperatures, given its lack of weight loss and low rate of change in weight, indicating its stability. However, after temperatures of 300°C, SPEEK retains more thermodynamic stability, and Nafion begins to decompose more rapidly.

**Proton Conductivity**

The proton conductivities at increasing temperatures must also be analyzed. The proton conductivity of a polymer is the ability of the polymer to transport hydrogen cations from the anode to the cathode. The faster this occurs, the faster more hydrogen can enter the initial anode area. If proton conductivity is slow, then, kinetically, it will be more difficult for the hydrogen to oxidize and create protons and electrons. Proton conductivity in itself is temperature dependent, and, based on the molecular structure of the membrane, temperature can further exasperate the rate of it.

Proton conductivities can be measured in a myriad of ways. Direct data that compares proton conductivity to temperature, for example, can be utilized. However, despite its importance, more analysis would be needed. One way to create a linear correlation between temperature and proton conductivity employs the Arrhenius equation (proton conductivity). This equation provides a linear relationship and insight into the activation energy and efficiency of proton conduction.

**Converting Arrhenius Equation**

\[
\sigma = \sigma_0 \times e^{-\frac{E_a}{RT}}
\]

Where
\(
\sigma = \text{Proton Conductivity (s/cm)}
\)
\(
\sigma_0 = \text{Pre Exponential Factor}
\)
\(
E_a = \text{Activation Energy (J/mol)}
\)
\(
R = \text{Universal Gas Constant (J/(mol*K)) = 8.3145 (J/(mol*K))}
\)
\(
T = \text{Temperature (K)}
\)

The natural log of each side can be taken.

\[
\ln(\sigma) = \ln(e^{-\frac{E_a}{RT}}) + \ln(\sigma_0)
\]
\[
\ln(\sigma) = -\frac{E_a}{RT} \ln(e) + \ln(\sigma_0)
\]
\[ \ln(\sigma) = -\frac{E_a}{R} \cdot \frac{1}{T} + \ln(\sigma_0) \]

Taking the natural log and graphing it against the reciprocal of temperature produce a linear graph, all else constant. As this is a line, a direct relationship between proton conductivity and temperature is apparent. The activation energy will also be expressed as the slope of this line.

**Proton Conductivity Data**

\[ \ast \sigma = \text{Proton Conductivity} \]

<table>
<thead>
<tr>
<th>Nafion</th>
<th>SPEEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>( \sigma \pm 0.01 )</td>
</tr>
<tr>
<td>( \pm 5^\circ \text{C} )</td>
<td>( 20 ) 0.0402</td>
</tr>
<tr>
<td>20</td>
<td>0.0402</td>
</tr>
<tr>
<td>30</td>
<td>0.0483</td>
</tr>
<tr>
<td>40</td>
<td>0.0562</td>
</tr>
<tr>
<td>50</td>
<td>0.0633</td>
</tr>
<tr>
<td>60</td>
<td>0.0704</td>
</tr>
<tr>
<td>70</td>
<td>0.0771</td>
</tr>
<tr>
<td>80</td>
<td>0.0841</td>
</tr>
<tr>
<td>90</td>
<td>0.0866</td>
</tr>
<tr>
<td>100</td>
<td>0.0878</td>
</tr>
</tbody>
</table>

With this data, two calculations need to be tabulated. First, the reciprocal of the temperature (in Kelvin) multiplied by 1000 is required, as this will provide the \( x \) values for the Arrhenius plot. Next, the natural logarithm of the proton conductivity will be taken. Blue Dagra software is used to analyze the literature data graphs. The uncertainties will be provided from the literature data and any appropriate calculations that may affect those uncertainties. Nafion was procured from the paper: “Preparation of Poly(Styrenesulfonic Acid) Grafted Nafion with a Nafion-Initiated Atom Transfer Radical Polymerization for Proton Exchange Membranes” (Peng et al., 2017). SPEEK data was procured from “Preparation and Characterization of Carbon Molecular Sieve (CMS)/SPEEK Bilayer Membranes and SPEEK/Polyimide (PI) Blend Membranes for Direct Alcohols Fuel Cell (DAFC) Performance (Maab et al., 2013).
An Analysis of SPEEK as a Suitable Alternative for Nafion in a Proton

<table>
<thead>
<tr>
<th>Nafion</th>
<th>SPEEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1000/T(K)$</td>
<td>$\ln(\sigma)$</td>
</tr>
<tr>
<td>3.41</td>
<td>-3.21</td>
</tr>
<tr>
<td>3.3</td>
<td>-3.03</td>
</tr>
<tr>
<td>3.19</td>
<td>-2.88</td>
</tr>
<tr>
<td>3.09</td>
<td>-2.76</td>
</tr>
<tr>
<td>3</td>
<td>-2.65</td>
</tr>
<tr>
<td>2.91</td>
<td>-2.56</td>
</tr>
<tr>
<td>2.75</td>
<td>-2.45</td>
</tr>
<tr>
<td>2.68</td>
<td>-2.43</td>
</tr>
</tbody>
</table>

Nafion Arrhenius Graph
Equation: $Y = -1.0849253135942911x + 0.5573968352467722$

Simplified Equation: $Y = -1.08x + 0.557$

**SPEEK Arrhenius Graph**

![SPEEK Arrhenius Graph]

Equation: $Y = -0.9289304552103271x - 0.5769165373875017$

Simplified: $Y = -0.929x - 0.577$

**Proton Conductivity Data Analysis**

The Arrhenius plot of the proton conductivities showcases varying degrees of overall proton conductivity at different temperatures. In addition, it provides insight into the capabilities of Nafion and SPEEK as temperature starts to increase. Based solely on the overall line for both SPEEK and Nafion, SPEEK’s overall Arrhenius plot is lower than Nafion’s plot. We can conclude that Nafion has a higher proton conductivity in comparison to SPEEK.
However, the rate of changes also provides important information. Nafion’s rate of change is of \(-0.108\) compared to SPEEK’s slope of \(-0.929\). Both rates of change are decreasing because both Nafion and SPEEK operate efficiently in an excess of water. This outcome can be attributed to the Grotthuss mechanism. As temperature increases, water starts evaporating, disrupting the mechanism and preventing efficient proton conductivity. SPEEK’s rate of change is lower in comparison to that of Nafion, indicating that it is less dependent on water in comparison to Nafion as temperatures increase. While the temperature gets negatively impacted, SPEEK’s overall reliance on water is much less due to the fact that SPEEK is much more regular than Nafion. Its regular structure and aromatic rings allow the SPEEK molecules to be physically closer, indicating that the sulfonic acid groups are in much higher concentration per mole of SPEEK compared to Nafion. More sulfonic acid groups lead to higher overall conductivity and less reliance on water. Thus, as temperature rises and water volume starts to decrease in the membrane, Nafion showcases a stark decrease in proton conductivity, while SPEEK slowly decreases but maintains consistency overall.

**Conclusion**

Based on the thermal analysis and the proton conductivities, SPEEK showcased more stability at high temperatures compared to Nafion, while Nafion showcased stronger proton conductivities. However, despite Nafion’s higher proton conductivity, SPEEK showcases much less degradation, an overall lower activation energy, and a decrease in proton conductivity at higher temperatures. Nafion, however, experiences drastic decreases. SPEEK exhibits a higher level of proton conductivity and surpasses Nafion’s temperature limit of \(100^\circ \text{C}\) due to its densely packed sulfonic acid group. This structural characteristic enables SPEEK to maintain its proton conductivity without experiencing significant losses even at elevated temperatures. Nafion’s overall superiority is essentially surpassed by SPEEK. Thus, SPEEK is, to a good extent, a suitable alternative for Nafion if it is being utilized at temperatures higher than \(100^\circ \text{C}\) where Nafion starts to diminish in effectiveness.

Our analysis did not consider other factors, including relative humidity, ion-exchange capacity, and sulfonation levels. These are also important factors that can determine more information about proton conductivity. They will also require their own investigations, potentially shedding more insight into the effectiveness of SPEEK and its suitability as a replacement for Nafion in the PEM. Furthermore, combining SPEEK with another polymer that is effective in proton conduction should be analyzed further. Nafion itself is thermally
stable as demonstrated; however, combining it with another polymer to create
a thermally stable co-polymer with high proton conductivity could potentially
allow for an important comparison between the new copolymer and Nafion.
Additional factors that affect proton conductivity need to be considered for this
analysis.

References

fonated carbon nanofibers for design and fabrication of SPEEK composite proton
exchange membranes for direct methanol fuel cell application. ResearchGate. https://
www.researchgate.net/publication/314648271_Electrospun_multifunctional_sul-
fonated_carbon_nanofibers_for_design_and_fabrication_of_SPEEK_composite_
proton_exchange_membranes_for_direct_methanol_fuel_cell_application
Liu, B., & Guiver, M. D. (2012). Proton conductivity of aromatic polymers. In Solid State Pro-
ton Conductors (pp. 331–369). Wiley.
Maab, H., & Shishatskly, S. (2013). Preparation and characterization of carbon molecu-
lar sieve (CMS)/SPEEK bilayer membranes and SPEEK/polyimide (PI) blend mem-
researchgate.net/publication/260276154_Preparation_and_characterization_of_car-
bon_molecular_sieve_CMS_SPEEK_bilayer_membranes_and_SPEEK_polyimide_
PI_blend_membranes_for_direct_alcohols_fuel_cell_DAFC_performance
gov/eere/fuelcells/parts-fuel-cell
Peng, K.-J., Yang, Y.-H., Lee, C.-Y., & Lin, K.-C. (2017). Preparation of Poly(Styrenesulfonic
Acid) Grafted Nafion with a Nafion-Initiated Atom Transfer Radical Polymerization
org/paper/Preparation-of-Poly(Styrenesulfonic-Acid)-Grafted-Peng-Yang/
b1ce77213937b0735142bdc1868d8617191f487d
“Proton Conductivity.” (n.d.). In ScienceDirect Topics. Retrieved April 30, 2023, from
https://www.sciencedirect.com/topics/engineering/proton-conductivity
Zhou, Su, and Fengxiang Chen. “PEMFC System Modeling and Control.” Fuel Cell Engi-
proton conductivity in polyperfluorosulfonic acid membranes. Journal of Membrane
Science, 192(1–2), 169–179. https://doi.org/10.1016/S0376-7388(00)85014-0
Sun, H., Li, H., & Li, B. (2015). A molecular dynamic simulation of hydrated proton
transfer in perfluorosulfonate ionomer membranes (Nafion 117). Journal of Chemis-
Thompson, E. L. (n.d.). Electron Microscope Analysis of Low Temperature Damage in
workgroups/c

Unlocking the Power of Molecular Cloning: Revolutionizing Medical Microbiology Procedures

NEELABH DATTA

The revolutionary realm of molecular cloning, encompassing the creation of recombinant DNA molecules, has ignited a wave of progress within the life sciences. The advent of potent tools has facilitated the manipulation of DNA, resulting in an extraordinary surge in the versatility and breadth of applications in recombinant DNA technology. The once-complex task of cloning genes has now been simplified, triggering a veritable explosion of insights into gene functionality. This has been achieved through the seamless fusion of multiple DNA fragments or the utilization of interchangeable gene cassettes, culminating in a state of unparalleled agility and expediency. In the 1970s, when restriction endonucleases, enzymes that cut DNA molecules selectively, were discovered, molecular cloning technology has grown exponentially in application and intricacy, resulting in influential DNA manipulation tools. Recent decades have seen an explosion in our understanding of gene function due to the simplicity and efficiency of molecular cloning. It is expected that emerging technologies will offer superior potentials, such as stitching together multiple DNA fragments in under a few hours and transforming the resulting plasmids into bacteria, or the use of swappable genes, which can be easily moved between different paradigms, maximizing promptness and flexibility. It has been proved that cloning techniques provide a gold standard technique for polymicrobial infection, recombinant cytokines, antimicrobial peptides, epidemiology, and gene therapy due to the limitations of culture-based methods. Due to the molecular cloning technique, recombinant antigens are now being used to monitor patients against clinical infections. As a result of laboratory techniques that permit in vitro chemical synthesis of any DNA construct specified in silico, molecular cloning will likely undergo a paradigm shift in the coming future. As a result of these advances, DNA clones can be constructed faster and iteratively, which will speed up the growth of new vaccines, gene therapy vectors, and recombinant proteins. Here I present a detailed overview of the latest applications of molecular cloning techniques in medical microbiology.

Contact: Neelabh Datta <neelabhdatta@gmail.com>
Keywords
molecular cloning, medical microbiology, gene therapy, antimicrobial peptides, recombinant cytokines, epidemiology, bioterrorism, polymicrobial infection

Introduction

Have you ever wondered how tiny organisms like bacteria and viruses can wreak havoc on the human body? Medical microbiology holds the key to unlocking the mysteries of these microscopic villains. Since the dawn of germ theory of disease, medical microbiology has progressed considerably at the end of the last century. It was at this time that premature culture techniques were fruitful in quarantining many pathogens and labeling them with their corresponding causative diseases. Medical microbiology has become an essential component of healthcare and society, thanks to its rapid advancement in response to clinical demands. This field of science investigates the relationship between large and small organisms in both normal and disease conditions, as well as the development of disease and the appropriate treatments that result in complete clinical recovery. A wide range of testing methods are employed to accomplish this task (Savitskaia, 1993). The field of medical microbiology continues to rely heavily on culture-based methods for identifying culture organisms despite substantial developments in analysis methodologies. By identifying the useful microorganisms from those that are pathogenic, medical microbiology contributes to the health of the public and helps to control infectious disease epidemics. Medical microbiology has undergone significant changes in the past few decades, with new methods being developed to make the isolation and detection of various microorganisms more efficient. In medical microbiology laboratories, a range of different microscopy and culture techniques are typically used. One example of a technique that can be used to detect bacteria, fungi, parasites, or viruses in infected cells is indirect fluorescent technique. This technique uses polyclonal antibodies raised in animals like mice or horses, as well as monoclonal antibodies produced using hybridization technology (Sharma et al., 2014). Latex agglutination tests are used to detect particulate antigens, while enzyme immunoassay tests are employed for soluble antigens (House et al., 2005). When it comes to life-threatening infections, it is essential to have quick and accurate diagnostic tests to facilitate prompt antimicrobial therapy (Sharma et al., 2014). In recent times, molecular biology techniques have emerged as a more efficient and rapid way to perform microbiological diagnosis. These methods use nucleic acid probes in conjunction with Polymerase Chain Reaction (PCR) amplification techniques (Leven et al., 1996). There is a possibility that cloning techniques could be useful in clinical settings, as they are already utilized in different areas...
of our everyday lives. For instance, during the outbreak of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2), laboratories recognized the need for the swift development of diagnostic tests for emerging diseases that have a significant impact on society. It is vital to quickly transfer such technology to laboratories that perform routine diagnostics. A contemporary issue is that scientific reactions to evolving threats are far more prompt than governmental reactions, and new diagnostic tests for use outside research laboratories often take a long time to be sanctioned (Raoult et al., 2004).

**From Past to Present: The Evolution of Molecular Diagnostics**

Molecular biology has been a vital tool for the medical laboratory, with its roots dating back to Linus Pauling’s discovery of sickle cell anemia as a “molecular disease” in 1949 (Pauling et al., 1949). However, it was not until recombinant DNA technology in the early days of molecular biology that it became usable in medical diagnostics (Chehab, 1993). Molecular diagnostics grew from basic knowledge on the primary sequence of various genes, with DNA probes incorporating radioactive nucleotides allowing analysis via Southern blotting of genomic regions. This led to the concept and application of restriction fragment length polymorphism (RFLP), which helped track variant alleles in the human genome (Williams, 1989). In 1978, molecular diagnostics techniques were used to make the first prenatal diagnosis of α-thalassemia, and the use of RFLP to characterize sickle cell alleles set the foundation for the characterization of other genetic diseases and infectious diseases using molecular diagnostics platforms (Kan et al., 1978). The mid-1980s brought the development of PCR, which quickly became a staple of laboratory medicine with its ability to exponentially amplify a target sequence and identify known mutations or sequences within hours (Mullis et al., 1986). PCR also established the foundation for many variant detection schemes based on the amplification of DNA. Following the publication of the human genome draft sequence, the challenge to improve existing variant detection technologies to achieve robust, cost-effective, rapid, and high-throughput analysis of genomic variation moved to the forefront of molecular diagnostics. Real-time PCR and its numerous variations, DNA microarray-based genotyping and transcription profiling, microbiome sequencing, proteomics, pharmacogenomics, nutrigenomics, forensic medicine, and CRISPR/Cas9 genome editing all represent important and critical advances in the field (Hendrix and Rohde, 2021). Despite the explosion of diverse variant detection assays, DNA sequencing remains the gold standard for pathogen identification and surveillance, especially with breakthroughs in next-generation sequencing (NGS) technology (Hendrix and Rohde, 2021). However, the costs of initial investment and difficulties in standardization and
interpretation of ambiguous results continue to limit the use of NGS in clinical laboratories. Physicians and other healthcare professionals are now working with molecular diagnostics professionals to understand the basis of infectious disease pathology and when to use molecular diagnostics like NGS. One example is the use of 16S in-house assay sequencing to identify bacterial pathogens directly from tissue specimens when culture results are negative, but there is evidence of histopathologic pathogen damage (Hendrix and Rohde, 2021). In addition to PCR and sequencing, molecular cloning also played a crucial role in the development of molecular diagnostics techniques. By enabling the replication of specific DNA sequences, molecular cloning made it possible to generate large amounts of identical DNA fragments for further analysis. This technique is based on the use of restriction enzymes to cut DNA at specific recognition sites and then ligating these fragments into a plasmid vector for amplification and manipulation. The ability to produce recombinant proteins using molecular cloning revolutionized the field of medical research and led to the production of many important biological therapeutics, including insulin, growth hormone, and clotting factors. Molecular cloning also allowed for the creation of genetic probes that could be used to detect specific DNA sequences, such as those associated with infectious agents. This enabled the development of highly sensitive and specific diagnostic tests, such as the PCR-based tests that are widely used today for the detection of viruses like HIV and hepatitis C. Molecular cloning techniques have also facilitated the identification of new disease-causing genes and the development of gene therapy approaches for genetic diseases.

An Overview of Molecular Cloning

In the annals of scientific progress, the field of molecular cloning stands as a testament to human ingenuity and perseverance. Throughout history, researchers have endeavored to unlock the secrets of DNA, seeking ways to manipulate and understand its intricate structure. It is within this context that the technique of molecular cloning emerged, revolutionizing the study of genetics and paving the way for a multitude of biological and technological applications. Molecular cloning (a molecular diagnostics technique) refers to a technique that involves isolating a specific DNA sequence and then amplifying short regions of it in vitro. The roots of molecular cloning can be traced back to the mid-20th century, when scientists began unraveling the mysteries of DNA. It was during this time that the concept of isolating specific DNA sequences and amplifying them in vitro took shape. One commonly employed approach involved the use of restriction enzymes, which acted as molecular scissors, cutting DNA at specific sites. By digesting existing DNA fragments or targeting them through PCR,
researchers were able to generate short inserts of DNA, typically around 100 base pairs in length (Juliane and Lessard, 2013). These resulting short inserts can also be created as complementary single-stranded fragments that are then annealed to form a double-stranded fragment (Juliane and Lessard, 2013). Once the DNA of interest has been obtained, it can be inserted into a vector plasmid, which is a circular double-stranded DNA. These vectors are smaller versions of naturally occurring plasmids and contain features such as replication origins, drug resistance genes, and unique restriction sites that enable the insertion of DNA fragments (Juliane and Lessard, 2013). The multiple cloning sites of these vectors usually contain different restriction sites, making it easier to select the appropriate enzymes for a variety of inserts. Molecular cloning was typically used to amplify DNA fragments containing genes, but it can also be used to amplify any DNA sequence, including promoters, non-coding sequences, chemically synthesized oligonucleotides, or fragments of randomly generated DNA (Sharma et al., 2014).

There was a wide range of biological applications and technological applications that make use of this method, including the production of recombinant antigens, cytokines, and proteins (Nguyen et al., 2004). If DNA sequences were to be amplified and cloned in vitro and in vivo, they must be linked to primary sequence elements that are capable of directing their own replication and propagation in the desired target host in conjunction with the linked sequence (Lu et al., 2008). Thus, the inclusion of a host-specific origin of replication and a selectable marker became essential sequence elements. In addition, when selecting a cloning vector, researchers had to consider a number of other characteristics, including the ability to express proteins, tag them, and generate single-stranded RNA and DNA (Sharma et al., 2014). As the field progressed, researchers sought to refine and expedite the cloning process. Recombinase-based cloning emerged as a one-step reaction, allowing for high-throughput cloning by inserting a specific DNA fragment into a specific region of target DNA through the interchange of relevant DNA fragments (Copeland et al., 2001). This streamlined approach facilitated the cloning of any DNA fragment, representing a significant leap forward compared to the classical restriction- and ligation-based approach, which involved fragmenting DNA with restriction endonucleases, ligating it to a vector, transfecting it into host cells, and subsequently screening and selecting the desired clones (Sharma et al., 2014). Cloning procedures generally followed these classical steps; however, a number of unconventional routes can be chosen depending on the specific solicitation.

Cloning procedures had traditionally adhered to the classical steps outlined previously, but the field of molecular cloning continued to evolve, offering researchers a growing repertoire of unconventional routes to choose from, based on the specific goals and requirements of their studies. Some advancements
had further expanded the possibilities and enhanced the efficiency of molecular cloning techniques. One notable development in molecular cloning was the emergence of advanced DNA assembly methods that enable the construction of large DNA constructs with precise control over their sequence composition. For instance, techniques such as Gibson assembly, Golden Gate assembly, and ligase cycling reaction (LCR) had revolutionized the process by allowing seamless assembly of multiple DNA fragments in a single reaction (Engler et al., 2008; Gibson et al., 2009; Li and Elledge, 2007). These methods bypassed the labor-intensive steps of traditional restriction digestion and ligation, streamlining the process and reducing the occurrence of unwanted mutations. Gibson assembly was a method that allowed the seamless assembly of multiple DNA fragments without the need for restriction enzymes or DNA ligases. It operated on the principle of in vitro homologous recombination and utilized three key enzymatic activities: exonuclease, polymerase, and DNA annealing. In this technique, the DNA fragments to be assembled were designed with overlapping regions called “homology arms.” These homology arms enabled the fragments to anneal to one another in the presence of the exonuclease and polymerase enzymes, which trimmed back the ends of the fragments and filled in the gaps, respectively. The resulting annealed fragments were then extended and ligated, producing a seamless composite DNA construct (Gibson et al., 2009). Gibson assembly offered several advantages, including its simplicity, efficiency, and the ability to seamlessly assemble multiple fragments with high fidelity, minimizing the introduction of errors. Golden Gate assembly was another powerful technique used for the modular assembly of DNA fragments. It relied on the activity of type IIS restriction enzymes, such as BsaI, which cut DNA sequences outside their recognition sites. The DNA fragments to be assembled were designed with specific recognition sequences for the type IIS enzyme at their ends, along with overlapping regions. During the assembly process, the type IIS restriction enzyme cuts the DNA fragments at the specific recognition sites, generating cohesive ends. The cohesive ends from different fragments were then ligated together, creating a composite DNA construct (Engler et al., 2008). Golden Gate assembly offered advantages such as modularity, versatility, and compatibility with high-throughput applications. It allowed the construction of complex DNA constructs by assembling multiple fragments in a single reaction, enabling researchers to efficiently generate diverse genetic constructs with precise control. LCR was a technique that combined the principles of PCR and DNA ligation to facilitate the assembly of DNA fragments. In LCR, short DNA oligonucleotides, called splints, were designed to anneal to specific regions on the DNA fragments to be ligated. These splints hybridized to their complementary sequences, bringing the DNA fragments in close proximity for ligation. The ligation step was mediated by a DNA ligase enzyme, which catalyzed the formation of phosphodiester bonds between
the adjacent DNA fragments. The ligated DNA fragments were then amplified by PCR using primers that hybridized to the ends of the assembled construct (Li and Elledge, 2007). These advanced molecular cloning techniques had transformed the field by offering streamlined processes for the assembly of multiple DNA fragments. They eliminated the need for laborious traditional methods, such as restriction digestion and ligation, and provided advantages such as increased efficiency, reduced errors, and the ability to construct complex DNA constructs with precision.

Another significant advancement was the utilization of site-specific recombinases, such as the Cre-Lox system and the Flp-FRT system, to enable precise DNA rearrangements. These recombinases catalyzed the exchange, inversion, or deletion of DNA segments at specific target sites, facilitating the generation of complex DNA constructs and genetic modifications (Sauer and Henderson, 1988; Buchholz et al., 1996). The Cre-Lox system, derived from the bacteriophage P1, consisted of two key components: the Cre recombinase enzyme and the LoxP sites. In the Cre-Lox system, the Cre recombinase recognized and bind to LoxP (locus of crossover [X] P1) sites present in the DNA. The LoxP site consisted of two 13-base-pair inverted repeats flanking a central 8-base-pair spacer region. The Cre recombinase binds to the LoxP site and catalyzes a strand cleavage and exchange reaction. This reaction involves the cleavage of the DNA strands at the specific LoxP sites, followed by the exchange or rearrangement of the cleaved DNA segments. The result was the excision, inversion, or rearrangement of DNA segments depending on the orientation and arrangement of the LoxP sites (Sauer and Henderson, 1988). The Cre-Lox system had been widely used for conditional gene knockout, gene expression control, and the generation of tissue-specific or inducible genetic modifications. This system offered precise control over DNA modifications, as the recombination events occurred exclusively at the LoxP sites and did not affect surrounding genomic regions. Similarly, the Flp-FRT system, derived from the yeast Saccharomyces cerevisiae, operated through the interaction between the Flp recombinase enzyme and the FRT (Flippase Recognition Target) sites. FRT sites are DNA sequences that serve as recognition sites for the Flp recombinase. The FRT site consisted of two 13-base-pair inverted repeats separated by a central 8-base-pair spacer region. The Flp recombinase recognized the FRT site and catalyzed a similar strand cleavage and exchange reaction as the Cre-Lox system. By cleaving the DNA strands at the FRT sites and facilitating the exchange or rearrangement of the cleaved segments, the Flp-FRT system enabled the excision, inversion, or rearrangement of DNA segments, depending on the arrangement and orientation of the FRT sites (Buchholz et al., 1996). The Flp-FRT system offered high precision and versatility in genetic manipulations, which allowed researchers to precisely control the outcome of the recombination events. The benefits of these site-specific recombinase systems are manifold.
First, they offered precise control over DNA modifications, allowing researchers to manipulate specific regions within a genome without affecting surrounding genetic elements. This targeted approach minimized unwanted off-target effects and preserved the integrity of the genome. Second, these systems provided flexibility and versatility in creating complex genetic modifications, such as gene knockouts, gene insertions, and conditional gene expression systems. The ability to precisely control DNA rearrangements at specific target sites opened up new avenues for functional studies and understanding gene regulation. Moreover, the reversible nature of these recombinase systems allowed for the creation of conditional genetic modifications, where DNA segments can be switched between different configurations, providing temporal and spatial control over gene expression.

In recent years, CRISPR-Cas9 technology has revolutionized molecular cloning and genetic engineering. Originally developed as a precise genome editing tool, CRISPR-Cas9 has been adapted for DNA cloning applications, allowing for efficient and targeted insertion of DNA fragments into specific genomic loci (Kleinstiver et al., 2014; Nishimasu et al., 2014). The CRISPR-Cas9 system has simplified and accelerated the process of creating transgenic organisms, generating knock-in or knock-out models, and facilitating the study of gene function (Doudna and Charpentier, 2014). The working principle of the CRISPR-Cas9 system involves two main components: the Cas9 enzyme and a guide RNA (gRNA). The Cas9 enzyme acts as a molecular scissor and can be programmed to target specific DNA sequences with the help of the gRNA. The gRNA is designed to recognize a complementary DNA sequence adjacent to the target site, guiding the Cas9 enzyme to that specific location. Once the Cas9 enzyme binds to the target DNA sequence, it generates a double-strand break (DSB) at that site. This DSB triggers the cellular DNA repair machinery, which can be harnessed to introduce desired DNA fragments into the genome (Doudna and Charpentier, 2014). CRISPR-Cas9 technology offers several significant benefits in molecular cloning and genetic engineering. First and foremost, it provides unparalleled precision in DNA targeting. The gRNA can be easily customized to recognize any desired DNA sequence, allowing researchers to selectively edit or insert DNA fragments at specific genomic loci. This level of precision enables the creation of highly specific genetic modifications, facilitating the study of gene function and the elucidation of molecular mechanisms. Another advantage of CRISPR-Cas9 technology is its efficiency and versatility. The Cas9 enzyme, guided by the gRNA, efficiently generates DSBs at the target site, triggering the cellular repair mechanisms. Researchers can exploit these mechanisms to introduce exogenous DNA fragments into the genome. This enables the precise insertion of DNA fragments, such as genes, regulatory elements, or reporter constructs, at desired genomic locations. In addition, the system can be used for gene knockouts by introducing
DSBs that lead to gene disruptions or deletions. CRISPR-Cas9 technology has significantly streamlined the process of creating transgenic organisms and generating genetically modified cell lines, providing researchers with a powerful tool for studying gene functions and disease mechanisms. The advancements in synthetic biology have opened new avenues for molecular cloning. DNA synthesis technologies have improved significantly, enabling the de novo assembly of entire genes, pathways, or even genomes (Kosuri and Church, 2014). This has paved the way for the creation of synthetic DNA constructs with tailored functions, such as metabolic pathways for bioproduction, novel enzymes, or even synthetic organisms with redesigned genomes (Gibson et al., 2010; Hutchison et al., 2016). These recent developments in molecular cloning techniques have not only increased efficiency but also expanded the possibilities for genetic manipulation, DNA engineering, and biotechnological applications. By embracing these cutting-edge approaches, researchers can push the boundaries of what is achievable in the field of molecular cloning.

The Role of Molecular Cloning in Polymicrobial Infections

Polymicrobial infections, where multiple microorganisms are involved in an infection, are a common occurrence in clinical settings. Molecular cloning had become an important tool in identifying diseases caused by multiple microorganisms and understanding the exchanges that occur within microbial communities. These diseases, as described by Kim et al. (2005), can be acute or chronic and are caused by combinations of viruses, bacteria, fungi, and parasites (Sharma et al., 2014). When a single pathogenic microorganism creates a niche, other pathogenic microorganisms can inhabit it, leading to colonization or the emergence of disease by two or more non-pathogenic microorganisms (Kim et al., 2005). Culture-based routine diagnostic testing is one treatment strategy, but it has limitations since it may not isolate all significant microbe species present in a sample. New bacterial community profiling techniques have revealed a greater diversity of microbes in infections caused by these bacteria than previously thought (Sharma et al., 2014). It is the result of these findings that polymicrobial infections are increasingly being perceived as complex communities of interacting organisms, whose pathogenicity is determined by dynamic processes (Rogers et al., 2010). One of the primary applications of molecular cloning in polymicrobial infections is the identification and characterization of individual pathogens. Traditional microbiological methods involve culturing the microorganisms from clinical samples, which can be time-consuming and may not always yield accurate results. In contrast, molecular cloning techniques such as PCR and NGS can identify and differentiate multiple pathogens in a single
sample rapidly and accurately. These techniques can identify pathogens that are difficult or impossible to culture, making them a valuable tool in the diagnosis of polymicrobial infections. However, even these techniques have certain drawbacks (Datta, 2023).

In the past few decades, advances in cloning and NGS technology had provided us with opportunities for gaining such insights about viable microbial cells (Wheat, 2010). An approach based on sequence homology allowed identification of bacteria in polymicrobial infections by cloning and sequencing the 16S ribosomal gene (Amann et al., 1995). With this method, it was possible to identify bacteria that died in the course of transportation or due to antibiotic treatment and to discover bacteria with specific growth requirements (Kommedal et al., 2009). However, it may not be possible to detect rare members of a community with divergent target sequences with rRNA gene-based cloning and characterization (Petrosino et al., 2009). Some of the limitations for it are due to primer bias and low sampling depth, which could be solved by 454 sequencing, pyrosequencing, or whole genome shotgun sequencing (Petrosino et al., 2009). Hence, molecular cloning techniques, such as Loop-mediated isothermal amplification (LAMP), Multi-Plex PCR-Based Reverse Line Blot Hybridization (mPCR-RLB), Target Enriched-Multiplex PCR (Tem-PCR), Gene chip technology, and Multiplex Ligation-Dependent Probe Amplification (MLPA), are utilized, which play a crucial role in identifying the genes and pathways involved in polymicrobial infections (Datta, 2023). These techniques enable researchers to gain a comprehensive understanding of the mechanisms underlying the interactions between pathogens, which is essential for developing novel treatments that target these interactions. LAMP is a powerful nucleic acid amplification technique that allows
for the detection and quantification of specific DNA sequences at a constant temperature. LAMP amplifies DNA by exploiting the properties of Bst polymerase and the unique structure of the target DNA. It involves the design of specific primers for different regions of the target gene, resulting in the formation of a loop structure. The amplification reaction produces a characteristic ladder-like structure (amplicon), which can be visualized using fluorescence or turbidity changes (Datta, 2023). LAMP is particularly useful for the prompt recognition of pathogenic microorganisms in laboratories with limited resources and experimental conditions. mPCR-RLB combines the sensitivity and the specificity of PCR with the high-throughput capabilities of reverse line blot hybridization. It involves the use of multiple primer sets that specifically bind to conserved regions of the microbial genome (Datta, 2023). These primers are used in a multiplex PCR reaction to amplify multiple target DNA sequences simultaneously. The amplified DNA is then hybridized to a reverse line blot membrane containing probes for different microbes. The presence or absence of hybridization signals allows for the identification of specific microbial types in the sample. Tem-PCR combines the specificity of PCR with target enrichment to detect and identify microbes involved in polymicrobial infections (Datta, 2023). The target enrichment step involves the selective amplification of specific regions of interest from the sample DNA using specific probes. This is followed by a multiplex PCR reaction, where multiple sets of primers amplify different targets in the same reaction. Tem-PCR allows for the simultaneous detection of multiple targets with high sensitivity and specificity. This technique is particularly useful for the detection of low-abundance targets in clinical samples. Gene chip technology, also known as DNA microarray, is a powerful tool for the detection and characterization of microbes. It allows for the simultaneous detection and analysis of multiple microbial genes, providing a comprehensive view of viral load and genetic diversity (Datta, 2023). Gene chips consist of a glass slide or silicon wafer coated with thousands of DNA probes complementary to specific regions of the microbial genome to be tested. When a sample of genomic DNA or cDNA is labeled and hybridized to the chip, the binding of target sequences to the complementary probes generates fluorescent complexes. The resulting signals provide information on the relative expression levels of the genes in the sample, enabling the identification of differentially expressed genes and grouping of samples based on their gene expression profiles. MLPA is a technique used to detect and quantify specific sequences in a sample. It combines the specificity of PCR with ligation and probe amplification. MLPA involves the ligation of two probes, one specific for the target sequence and the other for a control sequence, followed by PCR amplification using universal primers (Datta, 2023). The resulting pool of amplified fragments, proportional to the amount of DNA in the sample, can be quantified and analyzed to determine relative copy numbers of target sequences.
Molecular cloning also plays a crucial role in understanding the interactions between multiple pathogens in polymicrobial infections. For example, studies have shown that certain bacterial species can facilitate the growth of other pathogens, which can lead to more severe infections (Doron and Gorbach, 2008). Molecular cloning techniques can help identify the genes and pathways involved in these interactions, allowing researchers to better understand the mechanisms behind polymicrobial infections. This understanding can lead to the development of new treatments that target the interactions between pathogens rather than just targeting individual pathogens. Once the key genes and pathways responsible for pathogen interactions have been elucidated through molecular cloning techniques, researchers can delve deeper into understanding the underlying mechanisms. Armed with this knowledge, they can develop innovative treatments specifically designed to disrupt the cooperative or synergistic relationships between the pathogens involved. Several novel treatment strategies have emerged from this approach, each targeting different aspects of pathogen interactions. One such strategy involves interfering with quorum sensing, a cell-to-cell communication system employed by many bacteria to coordinate their activities (Rutherford and Bassler, 2012). Quorum sensing is a cell-to-cell communication system utilized by many bacteria, enabling them to synchronize their gene expression and coordinate their activities as a group (Rutherford and Bassler, 2012). This coordinated behavior often leads to the production of virulence factors, biofilm formation, or the regulation of key processes required for the survival and colonization within the host. Researchers have been able to identify the genes and molecules involved in quorum sensing through molecular cloning techniques, and by understanding the specific components of the quorum sensing system, they can develop therapeutic agents that interfere with this communication process and disrupt the signaling and coordination among pathogens (Vadakkan et al., 2018). There are different strategies employed to interfere with quorum sensing. One approach involves blocking or inhibiting the production or activity of quorum sensing signaling molecules, such as autoinducers (Vadakkan, 2020). These signaling molecules are produced by bacteria and accumulate in the environment as the population density increases and when their concentration reaches a threshold, they trigger specific gene expression programs, coordinating behaviors among the bacterial community (Vadakkan, 2020). Therapeutic agents targeting quorum sensing can be designed to mimic or block the action of these signaling molecules (Milly and Tal-Gan, 2023). For example, small molecules can be developed that either competitively bind to the receptor sites of bacteria, preventing the binding of natural signaling molecules, or mimic the signaling molecules themselves, leading to false or ineffective coordination among the pathogens (Milly and Tal-Gan, 2023). Another strategy involves inhibiting the production or activity of
the enzymes responsible for synthesizing the signaling molecules, which results in the disruption in communication and coordination among pathogens (Paluch et al., 2020). This interference can be achieved through the development of enzymatic inhibitors or by targeting the genes encoding the enzymes involved in the synthesis process (Paluch et al., 2020). Such an approach holds promising potential in combating pathogenic infections by impeding their ability to communicate and coordinate their detrimental actions.

The Role of Molecular Cloning in Antimicrobial Peptides

In order to combat the rapidly increasing antimicrobial resistance, more effective antimicrobial peptides are being developed. These antimicrobial peptides will kill target cells promptly and be effective against antibiotic-resistant and clinically pertinent pathogens. Advances in molecular genetics of antibiotic biosynthesis offer new opportunities to improve antibiotic production. The DNA of antibiotic makers is enriched with genes coding for antibiotic biosynthesis enzymes. There have been two types of vectors developed to clone antibiotic genes: low-copy and high-copy plasmids (Fakruddin et al., 2013). Many manufacturers are actively registering molecular clones of antibiotic synthesis genes. By manipulating the cloned genes encoding the enzymes involved in the biosynthetic pathway, new antibiotics can be produced as structural deviations of existing ones (Muller et al., 2007). These structural variants have diverse spectra and potency of activity against innumerable bacteria (Sharma et al., 2014). Moreover, the production of highly purified antimicrobial peptides at competitive costs is one of the biggest challenges facing the development of antimicrobial peptides.

Chemical peptide synthesis can be used to yield either native or modified cationic peptides, but it is more expensive than isolating peptides from natural sources. In vivo synthesis in host cells using recombinant technology is a more effectual and cost-effective method of synthesis (Muller et al., 2007). Recombinant expression of antimicrobial peptides is a promising area of research despite the challenges posed by their toxicity to bacteria and susceptibility to degradation by proteases (Sharma et al., 2014). This technology is widely recognized as an effective method for enhancing protein, peptide, and enzyme production. It offers advantages such as reduced time, well-established protocols, cost-effectiveness, and scalability (Sinha and Shukla, 2019). Bacteria and yeast are the most commonly used host systems for expressing recombinant products, including antimicrobial peptides (AMPs) (Gupta and Shukla, 2017). *Escherichia coli*, particularly the strain E. coli BL21 (DE3), is a popular choice due to its fast growth rate, high yields, established
expression protocols, and commercial availability of expression vectors. Other bacterial systems like *Bacillus subtilis* have also been used although to a lesser extent. Among yeast, *Pichia pastoris* has been employed as a potential host. Scientists have developed strategies to address these issues, such as fusing peptide genes with larger proteins and then using enzymes or chemicals to cleave them and release active peptides (Kuhn et al., 2003). Promoter probe vectors are also being employed to clone DNA sequences with transcriptional control signals. One interesting approach involves neutralizing the positive charge of the peptide to enhance its activity (Ali and Murrell, 2009). Another promising application of molecular cloning is the amplification of genes involved in biosynthesis pathways. By increasing the production of limiting enzymes, researchers have shown that it is possible to boost antibiotic production. This represents a significant step forward in the fight against antibiotic-resistant bacteria, which pose a major threat to public health. With the help of molecular cloning techniques, scientists are now better equipped to identify and develop novel antimicrobial agents that can be used to combat infectious diseases. Recombinant expression of AMPs is a popular method to produce large quantities of these peptides for further study or therapeutic use. However, the toxicity of some AMPs against bacteria and their susceptibility to proteolytic degradation can present challenges in the expression process (Nizet, 2006). To overcome these issues, researchers have used molecular cloning techniques to fuse the AMP genes with larger proteins. The fusion proteins can then be cleaved by enzymes or chemicals to release active AMPs (Ingham and Moore, 2007). To mitigate the toxicity of AMPs to the host strain, fusion proteins consisting of the target AMP and a carrier protein are often used. Carrier proteins possess anionic properties that neutralize the cationic charge of AMPs, reducing their toxicity and increasing solubility (Li, 2009). Common fusion partners include thioredoxin, Small Ubiquitin-like Modifier (SUMO), Glutathione S-transferase (GST), Biotin Carboxyl Carrier Protein (BCCP), and Green Fluorescent Protein (GFP). Cleaving the carrier protein from the target AMP typically requires the use of chemicals or enzymes. Chemical cleavage, although less specific, is considered more efficient than enzymatic cleavage, but it can introduce modifications in the side chain (Sinha and Shukla, 2019). Cleavage processes often leave one or two non-native residues at the N-terminus of the cleaved AMP (Sinha and Shukla, 2019). Affinity tags are sometimes conjugated to fusion partners to facilitate easy purification using affinity chromatography methods (Meiyagalan et al., 2014). Examples include His6-thioredoxin tagged GSL1 fusion protein, which can be purified by affinity chromatography, and fusion partners like glutathione S-transferase, which have inherent affinity properties and eliminate the need for additional tags (Schäfer et al., 2015). Self-cleaving tags with inducible proteolytic activity have also been

*UMURJ* • *vol. 17, no. 1 • 2024
used to simplify the separation of fusion tags. Thioredoxin and SUMO are the preferred fusion partners for the expression of recombinant AMPs (Xia et al., 2013). For instance, cecropin XJ, an insect AMP, was highly expressed in *E. coli* as a fusion peptide along with thioredoxin (Sinha and Shukla, 2019). Similarly, LsGRP1C protein production was assisted by a yeast SUMO tag in the *E. coli* host system, resulting in a high yield of the soluble SUMO-LsGRP1C fusion protein (Lin et al., 2017). BCCP has been used as a fusion protein for AMP expression, and *B. subtilis* has been employed for the recombinant expression of cathelicidin-BF from snake venom (Luan et al., 2014, Orrapin and Intorasoot, 2014). Another method for cloning AMPs is through the use of promoter probe vectors. These vectors contain transcriptional control signals that can be used to clone DNA sequences with transcriptional control signals by neutralizing the positive charge of the peptide (Muriana and Klaenhammer, 1991). This allows for more efficient expression of AMPs in various hosts. Furthermore, molecular cloning has also been used to amplify genes coding for limiting enzymes in biosynthesis pathways, which has been shown to increase the production of antimicrobial compounds (Adrio and Demain, 2010). This approach has led to the discovery and production of new antimicrobial compounds with potent activity against a number of microorganisms. Yeasts, particularly the methylotrophic yeast *Pichia pastoris*, have become increasingly important in genetic engineering and recombinant protein production due to their ease of genetic manipulation, ability to perform complex post-translational modifications, and rapid growth in cost-effective media (Kim et al., 2015). *P. pastoris* is a popular host for heterologous expression of recombinant AMPs (Ahmad et al., 2014). It offers advantages over *E. coli*, including the presence of a methanol-induced alcohol oxidase promoter, absence of endotoxins, correct protein folding capability, and suitability for large-scale production (Ahmad et al., 2014). In one study, the NZ17074 gene was synthesized and fused with SUMO3 in *P. pastoris* X-33, and the carrier protein was subsequently cleaved using formic acid (Wang et al., 2014). AMPs from plants, fruits, and chicken have also been expressed in *P. pastoris* but without the use of fusion proteins (Meng et al., 2017). The emergence of advanced gene editing tools, including Zinc-Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR-associated protein (CRISPR-Cas), has opened up new possibilities in the field of gene editing (Dangi et al., 2018). These tools make it easier to manipulate the genomes of expression hosts, enabling targeted modifications of specific genes to achieve desired outcomes. By harnessing gene editing technology, it is now possible to revolutionize the production of AMPs, particularly in response to the increasing demand for industrially and therapeutically valuable AMPs (Sinha and Shukla, 2019).
The Role of Molecular Cloning in Recombinant Cytokines

There are several fundamental homeostatic mechanisms that are modulated by cytokines (Huang et al., 2005). They include fever, acute phase infections, wound healing, inflammation, immune responses at the cellular levels, and tumor deterioration. Recombinant DNA technology has allowed us to clone the genes that encode these proteins, making it possible to use unrestrained measures of these cytokines to treat disease. There can be changes in amino acid sequence, absence of glycosylation (E. coli), and changes in glycosylation pattern (yeast, mammals, and insects) (Bandaranayake et al., 2011). Proteins expressed in the mature form in different host cells can also differ in their specific activities for several reasons (Sharma et al., 2014). Studies show that cytokines expressed in different host cells can have different pharmacokinetics, biologic properties, and immunogenicity due to physiochemical differences (Descotes, 2009). To explore the structure-function relationship of cytokines, expression vectors can be used to hypothesize recombinant forms. In addition to cytokines engineered for improved clinical efficiency or novel specificities, heterologous expression systems have also been used to create streamlined cytokines (Bermúdez-Humarán et al., 2011). Nowadays, recombinant cytokines are available as therapeutic agents. There has been evidence that GM-GSF or G-CSF can reduce the period and risk of infectious complications associated with chemotherapy-induced neutropenia (Sharma et al., 2014). Molecular cloning techniques are utilized to produce recombinant cytokines and enable their mass production. One of the most commonly produced cytokines through molecular cloning is interferon-alpha (IFN-α), which has potent antiviral and antiproliferative properties (Kumar et al., 2018). Recombinant IFN-α has been used in the treatment of hepatitis B and C infections, as well as certain types of cancer, including melanoma.
and leukemia. Molecular cloning is also used to produce other cytokines, such as interleukins and tumor necrosis factor (TNF). Recombinant TNF has been used in the treatment of bladder cancer, while interleukins have been used to enhance the immune response against cancer cells (Muller et al., 2003). In addition, the use of cytokines produced through molecular cloning has been explored in the treatment of autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis. Moreover, the investigation into the therapeutic potential of molecularly cloned cytokines has extended to encompass the management of autoimmune disorders, including rheumatoid arthritis and multiple sclerosis. These chronic conditions, such as rheumatoid arthritis, are characterized by persistent joint damage caused by an overactive immune response targeting the synovial membrane, cartilage, and bone (Zhang, 2021). While the precise cause of rheumatoid arthritis remains elusive, significant progress has been made in understanding its complex pathogenesis involving various cell types and signaling pathways (Verhoef et al., 2019).

Autoimmune processes and cytokines have emerged as key players in the initiation and perpetuation of rheumatoid arthritis. Notably, the extensively studied tumor necrosis factor (TNF), interleukin-6 (IL-6), and interleukin-1 (IL-1) have been implicated in the disease progression (Upchurch and Kay, 2012). Consequently, therapeutic strategies have focused on combating inflammation, with non-steroidal anti-inflammatory drugs and glucocorticoids serving as primary symptomatic management options (Zhang, 2021). In recent years, disease-modifying antirheumatic drugs (DMARDs) have revolutionized rheumatoid arthritis treatment by specifically targeting pro-inflammatory cytokines and their respective receptors (O’Dell, 2004). These biological DMARDs include TNF-α inhibitors (such as infliximab, etanercept, adalimumab, golimumab, and certolizumab pegol), IL-1 inhibitors, monoclonal antibodies against the IL-6 receptor (tocilizumab), T cell signaling inhibitors (abatacept), and monoclonal antibodies targeting CD20 (rituximab) (Klarenbeek et al., 2011). When combined with conventional synthetic DMARDs, early initiation of these treatments has shown promising results, leading to improved clinical outcomes and reduced joint damage (Zhang, 2021).

Along with using recombinant immunomodulators to enhance the host’s defense mechanism, attempts have been made to repackage defective immune genes (Miyake et al., 2001). The gene therapy of adenosine deaminase deficiency (ADA) deficiency has successfully treated patients with a defective T cell immune response (Aiuti and Roncarolo, 2009). It is possible to transfer a cloned ADA gene into lymphocytes of a patient using a retroviral vector; enzymatic and immune functions are thus reinstated (Aiuti and Roncarolo, 2009). Many ongoing research projects focus on gene therapy’s use in HIV infection and oncogenic virus-associated cancer, even though there is no clear evidence of its impact on infectious disease. Through HBV- or HHV-8-encoded surface receptors, the
thymidine kinase gene has been inserted into the cancer cell in HBV-associated hepatocellular carcinomas and HHV68-associated Kaposi’s sarcomas (Sharma et al., 2014). The patient is given ganciclovir when the thymidine kinase gene is integrated into their cancer cells, where the drug amasses and becomes toxic (Kieback et al., 2008). Future developments in the creation of cytokine receptor agonists show promise in a number of ways. One avenue of exploration involves using de novo protein design in addition to the existing combinatorial ligand engineering strategies. De novo design involves a computational approach to design analogues of interleukin-2 (IL-2) and interleukin-15 (IL-15) that can signal independently of IL-2Rα/IL-15Rα (Saxton et al., 2023). These synthetic analogues, known as “neoleukins,” have unique structural features compared to natural cytokines. They exhibit enhanced stability and improved effectiveness in mouse tumor models. Currently, neoleukins are undergoing phase I clinical trials for multiple cancer types, either alone or in combination with immune checkpoint blockade therapies. Advancements in artificial intelligence and machine learning-based protein structure prediction are expected to facilitate the development of even more sophisticated de novo designed cytokines (Saxton et al., 2023). These cytokines may have entirely distinct structural topologies that can modulate receptor geometry and composition in novel ways, potentially revolutionizing our understanding of cytokine signaling. While de novo cytokines offer advantages such as enhanced stability and potential ease of manufacturing, they may carry a risk of increased immunogenicity due to their non-human amino acid sequences (Saxton et al., 2023). However, ongoing efforts focus on minimizing immunogenicity by reducing the number of mutations and utilizing databases to assess potential neo-epitopes. Improvements in the pharmacokinetic and pharmacodynamic properties of cytokines are also crucial for their clinical success and strategies like half-life extension through Fc fusions or PEGylation, as well as local cytokine production via engineered T cells, aim to enhance cytokine safety and efficacy (Saxton et al., 2023). However, these modifications can influence cytokine signaling activity, necessitating a thorough understanding of their effects. Experimental models for screening and testing novel cytokine activities include in vitro cell culture and mouse models, but these have limitations in predicting human efficacy. Human patient-derived organoid cultures with diverse immune cell populations may provide valuable insights into the effectiveness of cytokine-based drugs in humans (Saxton et al., 2023). Immunogenicity is another significant consideration for protein therapeutics, including engineered cytokines. Non-human proteins, like de novo designed cytokines, may carry a higher risk of eliciting immune responses due to their divergent amino acid sequences, so efforts to minimize immunogenicity, especially for mutant cytokines, include reducing the number of mutations and monitoring for cross-neutralizing antibodies (Saxton et al., 2023).
The Role of Molecular Cloning in Gene Therapy

An individual can be treated with gene therapy by introducing a regular gene into their genome in order to mend the mutation that is the origin of genetic disease. In addition to possibly repairing the mutation, insertion of a regular gene into another functional gene may result in a new mutation if the normal gene integrates into another functional gene’s chromosomal site (Khan et al., 2016). Transformed cells may proliferate if normal genes replace mutant genes, leading to the restoration of the non-disease phenotype of the entire body. To date, human gene therapy has only been tested on somatic cells to treat cancer and severe immunodeficiency syndromes (Khan et al., 2016). It is possible for gene therapy to reverse the signs of disease in somatic cells, but the adjustment does not pass on to future generations. By placing corrected cells inside the germ line (e.g., cells of the ovary or testis), gene therapy will ensure that the next generation of cells undergoes meiosis and contributes to standard gametic development (Khan et al., 2016). In the field of health services, gene therapy is a progressive technique that has therapeutic prospect. The first successful report in gene therapy for the cure of genetic diseases gave physicians a promising approach to treating the lethal genetic disorders (Cavazzana-Calvo et al., 2000). Treatment for the primary immunodeficiency adenosine deaminase deficiency (ADA-SCID) shows good results with this approach. An improved gene transfer protocol and myeloablative conditioning regime, however, were later used to achieve fruitful results by aiming the hematopoietic stem cells (HSCs) (Aiuti et al., 2002).

The expression of specific genes by lentiviral vectors can correct X-linked disorders and adrenoleukodystrophy (X-ALD) (Cartier et al., 2009). Based on HIV-1 genes, X-ALD protein expression indicates gene-correction of true HSCs. As part of the treatment of metastatic melanoma through immunotherapy, lentiviral vectors were used for the first time in the cure of genetic human diseases

(a) Cytokine Receptor Activation: Here the common two-step activation seen in cytokine receptors is depicted. Initially, a soluble cytokine ligand (in yellow) binds to a high-affinity receptor subunit (in cyan) at “site 1.” This interaction recruits a low-affinity receptor subunit (in green) from elsewhere on the cell surface, forming a ternary cytokine receptor complex. This complex activates intracellular Janus kinases (JAKs) (in green and purple), leading to signal transducer and activator of transcription (STAT) transcription factor phosphorylation. (b) Diverse Cytokine Functions: Many cytokines exhibit functional and cellular diversity. This can result from multiple downstream transcription factors’ activation and their ability to signal in various cell types, such as macrophages (in cyan), T cells (in lilac), B cells (in red), and natural killer cells (in white). (c) Fine-Tuning Receptor Affinity: Modulating cytokine receptor affinity at “site 1” influences STAT activation’s dose sensitivity (EC_{50}). Changes at “site 2” affect the maximal strength (Emax) of this activation. Graphs depict how STAT phosphorylation (pSTAT) varies with cytokine concentration.
(Montini et al., 2012). By expanding the field of health sciences through immunotherapy, new opportunities were unlocked for treating serious diseases that cause death (Morgan et al., 2006). In two patients, continuous levels of T cells engineered to recognize tumors in the blood after infusion resulted in the recession of metastatic melanoma lesions up to a year after infusion. The bioengineered T cells were later studied for the treatment of chronic lymphocytic leukemia and metastatic synovial cell carcinoma where autologous T cells were innately altered to express Chimeric Antigen Receptors (CARs) that specifically bind to B cell antigen CD19 (Robbins et al., 2011). It has shown remarkable results for incurable autosomal recessive dystrophies, such as congenital blindness and Leber congenital amaurosis (LCA), in which gene transfer to a small number of cells at anatomically discrete sites has the potential to confer therapeutic benefit. Through gene therapy, a variety of cancers have been treated, including lung, gastrointestinal, hematological, gynecological, skin, urological, and neurological tumors (Khan et al., 2016). In order to treat diverse types of cancer, tumor suppressor genes have been inserted into immunotherapy, oncoturlytic virotherapy, and gene-directed enzyme prodrugs. In some cancer treatment strategies, p53 gene transfer is combined with chemotherapy or radiotherapy to increase the effectiveness of the tumor suppressor gene. An effective new anticancer agent (Ad5/35-EGFP) is being developed from fiber chimeric adenovirus vectors for the improved cure of hepatocellular carcinoma (Khan et al., 2016). In hepatocellular carcinoma (HCC), these vectors
were established to improve transduction and produce more virus progeny as a consequence of proper assaying. As a result of complex transgenic expression, in vitro HCC cells were found to possess enhanced antitumor activity while normal cells remained cytotoxic-free. As a result of the use of this technology, tumor growth was also repressed (Zhang et al., 2011). Recently, cancer gene therapy has gained more advanced technology and expanded its effectiveness (Lam et al., 2013). A mutation of the ABCA1 gene in high-density lipoproteins can cause the cells to discriminate into macrophages. Knockouts of this gene in embryonic stem cells augment the capability of these cells to differentiate into macrophages and precisely aim abnormal cells. A study of these allele replacements will provide insights into the regulatory variants that alter macrophage transcription and stability of mRNA (Smith, 2016).

Yan et al. (2020) have spearheaded a groundbreaking advancement in molecular biology by introducing a cutting-edge technique called “Nimble Cloning.” This pioneering method revolutionizes standardized molecular cloning, a fundamental technology in the field. By harnessing the combined power of the restriction enzyme Sfil and the T5 exonuclease, Nimble Cloning enables simultaneous vector linearization and generation of 3’-overhangs. Notably, this novel cloning system accommodates both PCR products and plasmids as inputs for the cloning reaction, rendering it highly efficient and adaptable to gene expression in both prokaryotic and eukaryotic systems. What sets Nimble Cloning apart is its remarkable versatility and simplicity. It empowers researchers to reuse DNA fragments or plasmid entry clones, facilitating efficient and streamlined workflows. This innovative method proves to be equally adept at cloning single or multiple fragments, as well as facilitating multi-site cloning. Consequently, the possibilities for modular assembly of DNA constructs are greatly expanded. In their groundbreaking research, Yan et al. (2020) introduced Nimble Cloning, a novel technique inspired by Gibson assembly, for the seamless assembly of DNA fragments. By leveraging the power of enzyme-catalyzed reactions, they demonstrated that Nimble Cloning surpasses traditional Gibson assembly in terms of cloning efficiency. Excitingly, this method eliminates the need for Taq DNA ligase, streamlining the process without compromising effectiveness. Through meticulous experimentation with single and multiple DNA fragments, Yan et al. successfully validated the superior performance of Nimble Cloning, consistently achieving over 99% positive clones. To further emphasize its versatility, they employed unique adapters in standardized cloning reactions, confirming that gene expression remained unaffected in both prokaryotic and plant systems. These findings not only establish Nimble Cloning as an efficient and reliable method for genetic engineering but also pave the way for future advancements in DNA fragment assembly.

In recent years, the CRISPR-Cas9 system has revolutionized the field of gene therapy. This revolutionary gene editing tool enables precise and efficient
Modification of specific DNA sequences within the genome. Molecular cloning plays a critical role in the construction of CRISPR-Cas9 vectors, which are used to deliver the Cas9 nuclease and guide RNA sequences into target cells. The ability to edit genes with unprecedented precision has opened up new possibilities for treating genetic diseases by correcting or modifying disease-causing mutations. Advancements in DNA synthesis technologies have facilitated the design and construction of synthetic genes and gene circuits for gene therapy applications. Molecular cloning techniques allow the assembly of these synthetic genes into expression vectors, enabling the production of therapeutic proteins or the regulation of gene expression in a controlled manner. In terms of gene delivery, tissue-specific targeting has become an area of active research, and by engineering viral and non-viral vectors to possess tissue-specific promoters or ligands, researchers aim to enhance the specificity and efficiency of gene delivery to target tissues or cells. This targeted approach minimizes off-target effects and improves the overall safety and efficacy of gene therapy. Innovations in genome engineering techniques, such as base editing and prime editing, have expanded the possibilities of gene therapy. These advanced tools enable precise modification of individual bases within the genome, offering the potential to correct disease-causing mutations without the need for introducing foreign DNA.

The Role of Molecular Cloning in Epidemiology

The rise of multidrug-resilient pathogens necessitates early molecular epidemiology outlining, both for comprehensive public-health reconnaissance and for timely treatment of infested patients. Due to the inoculum size and culturing conditions’ inconsistency, conservative tests of this type require extended culturing times (48–72 h) (Yang and Rothman, 2004). As genetic mechanisms of drug resistance are explained, nucleic-acid-based assays are being developed to report these inadequacies (Yang and Rothman, 2004). Three examples of how molecular epidemiology can be applied clinically are provided next. Although the absence of a resistance gene does establish a lack of resistance through that particular genetic mechanism, the presence of a resistance gene does not ineludibly infer its expression and conferment of phenotypic resistance: for example, the mecA gene is responsible for methicillin resistance (Yang and Rothman, 2004). With its high sensitivity and specificity, the mecA-PCR has become the most consistent method for detecting methicillin-resistant staphylococcus aureus (MRSA) due to its high detection sensitivity (Tenover et al., 1999) The detection of rifampicin resistance in M tuberculosis can also be performed using PCR-based resistance testing. RNA polymerase resistance in M tuberculosis is well characterized and is conferred by mutations within the DNA-directed RNA
polymerase subunit beta (rpoB) gene, which result in amino acid substitutions in the rpoB subunit (Telenti et al., 1997). The Line Probe assay (LiPA; Inno-Genetics) targets the mutation-prone rpoB gene segment (De Beenhouwer et al., 1995). A dramatic waning in detection time is provided to clinicians with this method, which is acute for treatment decisions, with an association of over 90% with typical resistance-detection methods (Marttila et al., 1999). There are many mutations and genetic loci involved in other M tuberculosis drug resistances, which makes genotyping more challenging (Yang and Rothman, 2004). Technical innovations like multiplex PCR or DNA microarray allow concurrent extension and analysis of multiple target sequences and will likely be able to overcome this future challenge (Elnifro et al., 2000). A PCR followed by nucleotide sequencing method is currently the most habitually used process to identify drug-resistant mutations in HIV-infected patients with ever-increasing indication supporting their prediction value (Demeter and Haubrich, 2001). Even though genotypic tests are more multifaceted than typical antimicrobial susceptibility tests, they provide perception into resistance development by perceiving mutations at concentrations too low to affect phenotypic assay susceptibility (Yang and Rothman, 2004). Also, they have the benefit of detecting mutations that do not cause drug resistance but do designate selective drug pressure, which could influence treatment decisions for individual patients.

A recent study by Zhang et al. (2021) showed that the pathogen Carbapenem-Resistant Acinetobacter baumannii (CRAB) poses a significant threat to both health and economy. It haunts hospitals, using them as breeding grounds for its insidious transmission. The vulnerable victims of CRAB’s infections are often those who find themselves confined within hospital walls, with open wounds and lengthy stays. The elderly, whose natural defenses have weakened, are especially susceptible. Strikingly, most patients in the study suffered from multiple diseases simultaneously, with pulmonary infections, such as pneumonia and respiratory failure, being the most common. This aligns with the well-known association between CRAB and pneumonia and bloodstream infections. The key to CRAB’s resistance lies in its ability to produce carbapenemases, such as blaOXA-23 and blaOXA66, which render it impervious to carbapenem antibiotics (Zhang et al., 2021). This study by Zhang et al. (2021) confirmed that all strains harbored these carbapenem resistance genes. The mechanisms behind CRAB’s increased antibiotic resistance are multifaceted, involving mobile genetic elements, chromosomal β-lactamases like blaADC, and the presence of efflux pumps (Zhang et al., 2021). The study also highlighted the presence of various genetic structures, including ISs and AbaR-type genomic resistance islands, which facilitate the spread of antibiotic resistance determinants among pathogens, compromising treatment efficacy. But CRAB is not merely resistant; it is also armed with virulence factors that allow it to thrive in the inhospitable environment of a
hospital. The study also revealed that all CRAB strains possessed a repertoire of virulence genes, such as \textit{bap}, \textit{csuABCD}, \textit{pgaABCD}, \textit{bfmRS}, \textit{entE}, \textit{ompA}, and \textit{plcD}. These genes contribute to the formation and maintenance of biofilms, which are strongly associated with MDR strains and clinical infections. In addition, CRAB isolates carried genes involved in the production and uptake of the acinetobactin siderophore, further enhancing their infectious prowess (Zhanget al., 2021). When it comes to the origins of CRAB outbreaks, molecular cloning holds the key to unraveling the mystery. Previous reports have linked clonal outbreaks to international clone lineages, with European clones I, II, and III being the most prominent. In the study conducted by Zhang et al. (2021), all CRAB isolates belonged to ST2, which is associated with outbreaks and harbors the \textit{blaOXA-23} gene. Interestingly, certain clone types, such as ST2, ST25, and ST78, exhibited enhanced biofilm formation, likely contributing to their success in colonizing the clinical environment (Zhanget al., 2021). Through the lens of core genome phylogenetic analysis, it became clear that specific CRAB strains dominated the scene. Cluster 1 and cluster 2 emerged as the primary players, with clone 1 and clone 2 persisting throughout the study period (Zhanget al., 2021). However, changes in clone groups were observed over the years, indicating the dynamic nature of CRAB populations. These clones, particularly clone 1, underwent population expansion and exhibited distinct resistance, virulence, and insertion sequence patterns compared to clone 2 (Zhanget al., 2021). These findings emphasize the genetic diversity within CRAB outbreaks and the importance of monitoring and controlling the spread of specific clones.

The Role of Molecular Cloning in Bioterrorism

Biological warfare, in its essence, is the strategic utilization of living organisms or their by-products to cause harm or destruction. It involves the deliberate release or deployment of pathogens, toxins, or biological agents with the intent to incapacitate, debilitate, or even kill living organisms, including humans. This form of warfare relies on exploiting the natural abilities of organisms to produce toxins or propagate diseases, harnessing their destructive potential as a means of achieving military or political objectives. Biological warfare encompasses a range of tactics, from the use of infectious diseases as weapons to the manipulation of toxic substances derived from a number of organisms. Its history intertwines with humanity’s understanding of the natural world and our capacity to wield its forces for destructive purposes. An anthrax outbreak that occurred has drawn much attention to the growing threat of bioterrorism. Responsibility of the clinicians will be crucial in instigating applicable response actions if they can recognize and diagnose real or suspected bioterrorism events quickly and
accurately (Pavlin et al., 2002). It is difficult to discriminate between a bioterrorism victim’s symptoms and symptoms of an ordinarily encountered disease process, as was the case at the time of the 2001 anthrax episode (Pavlin et al., 2002). In suspected clinical outbreaks, conventional culture-based assays cannot detect bioterrorism agents due to the formerly designated restrictions. As a result of the prolonged incubation required by conventional microbiological methods, the laboratory is exposed to increased biohazard risks because of the redundant proliferation of bioterrorism pathogens (Yang and Rothman, 2004). Many bioterrorism agents have, of late, been studied using PCR-based assays, including *Variola major*, *Bacillus anthracis*, *Yersinia pestis*, and *Francisella tularensis* (Espy et al., 2002). Bioterrorism agent PCR diagnostics was used for both screening preclinical victims for early prophylactic treatment and diagnosing symptomatic individuals (Yang and Rothman, 2004). Despite the similarity between most bioterrorism-induced illnesses and natural outbreaks, it was possible that the contributory agents of bioterrorism may have been genetically engineered to be more virulent, impervious to antibiotics or vaccines, or to produce phenotypic characteristics resembling multiple infections via the insertion of recombinant genes (Alibek, 1999). Since DNA-based methodologies are more easily adjustable and proficient in uncovering more comprehensive information embedded within genetic sequences, they are likely to be more valuable than conventional detection methods in such cases (Yang and Rothman, 2004).

The potential role of genetic engineering in enhancing the lethality of infectious agents is a topic often discussed in relation to biological warfare. While there is some validity to this claim, it is crucial to consider the following perspective: Imagine a scenario where a bioterrorist endeavors to genetically modify a harmless laboratory bacterium, such as *E. coli*. The aim would be to render the bacterium “invisible” to the human immune system upon entry into the body (Clark and Pazdernik, 2016). In addition, the bacteria could be engineered to release toxins, thwarting immune cells and introducing genes to impede the vital iron supply from blood cells. Lastly, the bacterium could be modified to possess high levels of infectivity. Such a biological agent would undoubtedly pose a formidable threat. Astonishingly, this bacterium already exists in nature—it is known as *Yersinia pestis*, the notorious causative agent of bubonic plague (Clark and Pazdernik, 2016). Endemic in different parts of the world, including China, India, Madagascar, and the United States, this pathogen demonstrates how nature has provided an exceptionally dangerous biological weapon. Consequently, the notion of enhancing infectious diseases through genetic engineering appears to be a minor concern. Although information regarding the Soviet germ warfare facility’s modification of the smallpox virus and the creation of artificial mutants and hybrids remains largely undisclosed, recent experiments involving mousepox (Ectromelia virus) have yielded alarming outcomes. Mousepox, a virus
primarily infecting mice, exhibits varying degrees of virulence depending on the mouse strain. Genetically resistant mice rely on cell-mediated immunity rather than antibodies to combat the virus, with natural killer (NK) cells and cytotoxic T cells effectively eliminating infected cells and clearing the virus from the body. In an attempt to improve and balance the immune response, researchers introduced the human gene for the cytokine interleukin-4 (IL-4) into the mousepox virus. IL-4 stimulates B cell division and antibody synthesis, which theoretically should have led to an enhanced immune response (Clark and Pazdernik, 2016). Surprisingly, the results were contrary to expectations, as the engineered virus displayed significantly heightened virulence. It not only caused mortality in all genetically resistant mice but also claimed the lives of 50% of vaccinated mice. Excessive IL-4 expression suppressed NK cells and cytotoxic T cells while failing to enhance the antibody response (Clark and Pazdernik, 2016). Similar results have been observed with different strains of Vaccinia virus, which is utilized for smallpox vaccination. The repercussions of inserting IL-4 or other immune regulators into smallpox itself and the potential to undermine the immune response and increase virulence remain uncertain (Clark and Pazdernik, 2016). Poxviruses possess genes called cytokine response modifier (crm) genes, designed to hinder the action of NK cells and cytotoxic T cells, further complicating the assessment of smallpox’s virulence (Clark and Pazdernik, 2016). Nevertheless, genetic engineering allows for the concealment of a potentially hazardous virus within a harmless bacterium, a phenomenon already witnessed in nature when bacteriophages incorporate their genomes into bacterial chromosomes or plasmids, subsequently re-emerging to infect other hosts. Theoretically, cloning the complete genome of a small animal or plant virus into a bacterial plasmid could serve as the basis for a biological weapon, while larger viruses could be accommodated using bacterial or yeast artificial chromosomes. In the case of RNA viruses, the viral genome must first be reverse transcribed into complementary DNA (cDNA) before cloning into a bacterial vector. Viruses harboring toxic sequences, which are not stably maintained on bacterial plasmids, could potentially be cloned as separate fragments. This approach has proven successful with yellow fever virus albeit requiring in vitro ligation of the fragments to generate a complete, functional cDNA (Clark and Pazdernik, 2016). Different cell types, including bacteria and eukaryotes, are capable of taking up DNA or RNA under specific conditions through transformation. Consequently, the naked nucleic acid genomes of numerous DNA and RNA viruses retain their infectivity even without their protein capsids or envelopes. Once a viral genome is cloned, the DNA molecule containing it becomes infectious itself. Alternatively, cDNA versions of RNA viruses can infect host cells, giving rise to new virus particles containing RNA. Poliovirus, influenza, and coronavirus are among the RNA viruses that have demonstrated this capability. An ingenious strategy for generating an
RNA virus involves cloning the cDNA version of its genome downstream of a strong promoter in a bacterial plasmid. Transcription within the induced bacterial cell results in the production of a large number of infectious viral particles. Combining a dangerous human RNA virus with a harmless intestinal bacterium, controlled by a promoter responsive to intestinal conditions, could pose a significant threat (Clark and Pazdernik, 2016). While certain pathogenic bacteria exhibit slow growth or are challenging to culture outside their host organisms, advancements in biotechnology have facilitated the identification of infectious microbes through molecular diagnostics. Instead of relying on classical microbiological techniques to grow and identify disease-causing agents, molecular diagnostics employ the analysis of molecules, primarily DNA but also RNA, proteins, and volatile organic compounds. These molecular techniques offer advantages in terms of speed, accuracy, and sensitivity.

Fluorescent in situ hybridization (FISH) is one such diagnostic method, involving the direct probing of biopsies or patient samples with fluorescent DNA oligonucleotides specific to a target pathogen (Clark and Pazdernik, 2016). If the pathogen is present, the probe binds to complementary DNA in its chromosome, enabling visualization of fluorescence under a microscope. An innovative approach utilizing peptide nucleic acid (PNA), which replaces the negatively charged sugar-phosphate backbone of DNA with a neutral peptide backbone, enhances the binding affinity of PNA probes to complementary DNA and facilitates their entry into bacterial cells. PCR amplification of target DNA sequences, unique to a particular pathogen, is another commonly used method in molecular diagnostics (Clark and Pazdernik, 2016). The ability to design primers specific to a pathogen allows for the amplification of target DNA, enabling PCR to serve as a diagnostic tool. PCR’s versatility stems from its potential to detect a single molecule of target DNA and its applicability to microbes that are challenging to culture in the laboratory. However, PCR is susceptible to contamination and false positives. Randomly amplified polymorphic DNA (RAPD), a PCR variant, distinguishes different strains of the same bacterial species, aiding in epidemiological studies to track the spread of infectious diseases. Each microorganism species possesses a unique small-subunit ribosomal RNA (SSU rRNA) sequence, such as 16S rRNA in bacteria and 18S rRNA in eukaryotes (Clark and Pazdernik, 2016). Therefore, clinicians can employ PCR to amplify the gene encoding the microbe’s SSU rRNA when faced with an unknown infection. By sequencing and comparing the PCR fragment to a database of known DNA sequences, the pathogen can be identified. Checkerboard hybridization is a technique that utilizes SSU rRNA as a basis, allowing for the simultaneous detection and identification of multiple bacteria in a single sample. Horizontal lines on a hybridization membrane contain probes specific to different bacteria, while PCR amplification of the SSU rRNA gene from clinical samples, potentially harboring a mixture of
Resounding Expression of Cloned RNA Viruses. Credits to Clark and Pazdernik, 2016
pathogens, generates fluorescently labeled fragments that are applied vertically to the membrane. Fluorescent spots indicate samples that have hybridized with the probes. Abbott Laboratories has developed a potentially revolutionary technology called PLEX-ID, which combines traditional PCR with mass spectrometry to identify unknown microbes in patient samples (Clark and Pazdernik, 2016). By analyzing the mass of amplified DNA fragments using a mass spectrometer, the DNA sequence can be deduced, enabling the identification of the pathogen. PLEX-ID has the capability to provide a diagnosis within eight hours. In the future, disease diagnosis may be achievable using an “electronic nose,” a device that detects volatile organic compounds released by pathogens or by the body in certain diseased conditions (Clark and Pazdernik, 2016).

The Role of Molecular Cloning in SARS-CoV-2 Treatment

B cell receptor (BCR) repertoires display remarkable sequence diversity resulting from somatic recombination and hypermutation processes occurring during B cell development. BCRs are transmembrane receptors situated on the surface of B cells, and their variable regions interact with specific antigen epitopes to initiate an immune response through antibody production (Zhou et al., 2021). This variable region shares identical gene sequences with the corresponding antibody produced by the B cell. The variability in the variable region is generated by somatic recombination of three gene segments in the heavy (H) chain locus (V, D, J) and two gene segments in the light (L) chain locus (V, J) (Zhou et al., 2021). The variable region of an antibody, also known as immunoglobulin (Ig), determines its specificity for binding to a specific viral antigenic epitope. Somatic hypermutation occurs during B cell proliferation in the germinal center, introducing random mutations in the genes encoding the variable region of individual monoclonal antibodies (mAbs) (Zhou et al., 2021). This process is crucial for the development of high-affinity antibodies, a phenomenon referred to as antibody affinity maturation. Importantly, no two B cells possess identical BCRs. Therefore, the cloning of a functional mAb requires the analysis and cloning of one B cell at a time to ensure the native pairing of antibody heavy and light chains (Bassing et al., 2002).

Advancements in antibody gene cloning techniques, such as hybridoma technology, human B cell immortalization, antibody phage display, human immunoglobulin transgenic mice, and single B cell antibody technology, have revolutionized the process of cloning functional human neutralizing antibodies (HuNAbs) (Tiller, 2011). Among these techniques, single B cell-based antibody cloning has been extensively employed to isolate SARS-CoV-2-specific antibodies during the ongoing COVID-19 pandemic. This method involves amplifying
auto-paired Ig heavy- and light-chain RNA sequences from a heterogeneous population of memory B cells, followed by their in vitro construction into functional mAbs (Niu et al., 2019). The successful application of this technique has been demonstrated in the isolation of neutralizing mAbs against various viral infections, such as human immunodeficiency virus (HIV) and SARS-CoV-2 (Corti et al., 2015). In recent years, the combination of single-cell reverse transcription polymerase chain reaction (RT-PCR) and single B cell sorting has significantly improved the success rate of antibody gene cloning. The acquisition of single antigen-binding memory B cells through fluorescence-activated cell sorting (FACS) or optofluidic platforms has been a major technical advancement (Zhou et al., 2021). These techniques enable subsequent nested RT-PCR using primers designed to amplify naturally paired antibody heavy- and light-chain gene sequences from individual memory B cells (Liao et al., 2009). Furthermore, high-throughput single-cell RNA and VDJ deep sequencing of BCR repertoires, coupled with bioinformatics analysis, have surpassed single-cell RT-PCR in terms of efficiently screening large pools of diverse memory B cells (Cao et al., 2020). Notably, immunization using RBD/S proteins or direct infection of mice with genetically humanized immune systems has been demonstrated to generate complete HuNAbs against SARS-CoV-2 (Hansen et al., 2020). This platform has successfully facilitated the cloning and screening of certain SARS-CoV-2-specific HuNAbs. Deep sequencing of variable regions and BCR repertoires has provided insights into the characteristics of human antibody heavy and light chain pairing (Zhou et al., 2021). In addition, the combination of microfluidic-based techniques and bioinformatics analysis has the potential to further enhance the efficiency of identifying highly potent HuNAbs against specific viral antigens (Hansen et al., 2020). These advancements hold implications for combating COVID-19 and other emerging infectious diseases.

Phage display has emerged as a widely utilized approach for cloning human antibodies. This technique involves two main stages: constructing an antibody gene library and screening for antibodies specific to a particular antigen (Zhou et al., 2021). To generate a human Fab library, researchers can extract peripheral blood mononuclear cells (PBMCs) from individuals who have recovered from COVID-19. By amplifying the antibody gene pool using specific primers targeting the variable regions of the heavy and light chains, a phage library can be constructed (Zhou et al., 2021). The library can also be synthesized using selected human germline immunoglobulin variable segments, with diversity introduced in the complementarity-determining regions (CDRs) through mutagenesis. The library can be screened using SARS-CoV-2 receptor-binding domain (RBD) as a bait to isolate RBD-specific human Fabs (Zhou et al., 2021). These Fabs can then be further developed into full-length IgG1 antibodies for subsequent testing of their biochemical and functional properties. Although the phage display
technique has some limitations, such as the unnatural pairing of heavy and light chains and a time-consuming panning procedure, it remains a valuable tool for obtaining antigen-specific antibodies, including potent SARS-CoV-2-specific human neutralizing antibodies (HuNAb) (Zhou et al., 2021). The SARS-CoV-2 spike protein (S protein) plays a crucial role in viral entry by binding to the ACE2 receptor on host cells (Lv et al., 2020). The receptor-binding domain (RBD) within the S1 subunit of the S protein undergoes conformational changes, transitioning between “up” and “down” states. The “up” conformation is particularly targeted by HuNAb since it represents the active state for ACE2 binding (Zhou et al., 2021). Multiple neutralizing domains within the S1 region have been identified, including RBD and the N-terminal domain (NTD). HuNAb can be classified into three main groups based on their binding sites (Yuan et al., 2021). The first group, known as receptor-binding site (RBS) antibodies, targets epitopes within the RBD that overlap with the ACE2 binding site (Brouwer et al., 2020). RBS antibodies can further be divided into subclasses based on their interaction angles with the viral RBD (Yuan et al., 2021). The second group includes cross-reactive antibodies, such as CR3022, which recognize cryptic sites in the RBD and exhibit varying neutralizing activities against both SARS-CoV and SARS-CoV-2 (Yuan et al., 2021). The third group is represented by antibodies like S309, which engage the RBD through an epitope containing the N343 glycan and can neutralize both SARS-CoV-2 and SARS-CoV (Yuan et al., 2021).

Several anti-SARS-CoV-2 HuNAb have been developed and undergone clinical trials. Paired HuNAb have been formulated to enhance the breadth of neutralization against emerging SARS-CoV-2 variants and minimize the occurrence of HuNAb escape variants. HuNAb combinations under Emergency Use Authorization (EUA) exhibit live virus neutralization IC50 values below 0.1 µg/ml (Jones et al., 2021). Typically, these paired antibodies consist of one RBS-A HuNAb along with either one RBS-B or one RBS-C HuNAb. For example, the combination of LY-CoV555 (bamlanivimab, BAM) and LY-CoV016 (etesevimab, ETE) from Eli Lilly has been administered to COVID-19 patients (Zhou et al., 2021). The randomized phase 2/3 trial demonstrated a statistically significant reduction in viral load at day 11 among non-hospitalized patients with mild to moderate illness who received the combination therapy (Gottlieb et al., 2021). Similarly, the REGN-CoV2 antibody cocktail (casirivimab, CAS, and imdevimab, IMD) showed a substantial reduction in viral load in patients with baseline viral load higher than 107 copies/mL compared to the placebo group (Hansen et al., 2020). Due to concerns regarding reduced neutralizing abilities against SARS-CoV-2 variants of concern (VOCs), ongoing clinical trials are investigating the efficacy of more broadly reactive HuNAb, as well as next-generation antibodies such as bispecific antibodies and engineered antibodies. In addition, the development of BRII-196 and BRII-198, an antibody combination approved for clinical
use in China, has shown promising results with 78% efficacy (Zhou et al., 2021). SARS-CoV-2 virus has the ability to undergo mutations during its replication in humans. Unlike the SARS virus in 2003, which was eliminated, COVID-19 has persisted, leading to concerns about viral mutations that could evade immunity from natural infection or vaccination (Romano et al., 2020). Several VOCs have emerged during the pandemic, including the Alpha, Beta, Gamma, Delta, and Omicron variants (Chakraborty et al., 2022). These VOCs have mutations in the spike (S) glycoprotein of the virus, which can confer resistance to human neutralizing antibodies (HuNAbs) and increase viral transmissibility. The D614G mutation, found in all VOCs, enhances viral transmissibility (Plante et al., 2021). Mutations such as E484K/A and amino acid deletions in the S protein reduce the potency of HuNAbs targeting specific regions (Jangra et al., 2021). The emergence of these VOCs poses challenges to public health and vaccine responses.

Efforts have been focused on developing vaccines to combat COVID-19, and several vaccines have been approved for emergency use. These include mRNA-based vaccines, adenovirus-vectored vaccines, and inactivated vaccines (Zhou et al., 2021). The efficacy rates reported for these vaccines mainly reflect their ability to prevent severe diseases, while their effectiveness against SARS-CoV-2 nasal infection and VOCs requires further investigation (Zhou et al., 2021). Studies have shown reduced neutralization of VOCs by antibodies from vaccinated individuals, indicating the potential for decreased effectiveness against certain variants (Wang et al., 2021). However, widespread vaccination campaigns have contributed to reduced hospitalizations, deaths, and infections in various countries. Future strategies for vaccination and immunotherapy should aim to provide broadly reactive immune protection against multiple variants (Zhou et al., 2021). The role of mucosal immunity and tissue-resident memory T cells in the upper respiratory tract needs to be explored for long-term protection (Baum et al., 2020). In addition, careful monitoring of antibody-mediated enhancement of SARS-CoV-2 infection and immunopathogenesis is crucial in the context of human infections and clinical management.

Synergy of Molecular Cloning and the CRISPR-Cas System in SARS-CoV-2 Treatment

COVID-19, caused by the SARS-CoV-2 virus, has triggered a global pandemic, with the emergence of highly transmissible and more fatal mutant strains. Effective management of this infectious disease requires a number of measures, including social distancing and isolation of confirmed cases. However, there is a pressing need for a highly sensitive diagnostic kit that enables rapid early detection (Lou et al., 2022). Diagnostic approaches for COVID-19 involve immunological
and molecular techniques. Immunological methods detect viral antigens or antibodies in the lungs or blood, aiding in disease understanding and transmission dynamics (Mahase, 2020). Molecular methods primarily focus on nucleic acid detection, with RT-PCR being the gold standard due to its high sensitivity and accuracy (Broughton et al., 2020). However, the limited availability of RT-PCR equipment and materials can lead to delays and false negative results, making it unsuitable for mass screening (Broughton et al., 2020, Li and Ren, 2020). Hence, there is a need for simplified and time-efficient diagnostic methods with high accuracy (Xiang et al., 2020). The CRISPR-Cas system, with nucleic acid detection technologies like SHERLOCK (Cas13a), DETECTR (Cas12a), CDetection (Cas12b), and CAS14-DETECTR, offers a new avenue for pathogen screening (Ding et al., 2021). In recent times, diverse detection methods combined with isothermal amplification and the CRISPR-Cas system have emerged as rapid diagnostic tools for detecting SARS-CoV-2 viral RNA. For instance, a visual analysis method called CLAP combining AUNP and Cas12a-assisted RT-LAMP demonstrated the ability to detect low levels of SARS-CoV-2 RNA rapidly (Lou et al., 2022). In addition, engineered Cas12a enzyme-based LAMP showed promising results in detecting wild-type and mutant SARS-CoV-2 within a short time frame (Lou et al., 2022). Cas13a crRNAs can specifically target SARS-CoV-2 and related coronaviruses, and when combined with a generic autonomous enzyme-free hybridization chain reaction (HCR), they offer a detection method for monitoring virus transmission via objects (Yang et al., 2021). The CRISPR-Cas13 amplification strategy holds potential for efficient surveillance of SARS-CoV-2 transmission (Lou et al., 2022). Viral control over host cells has long intrigued virologists due to the limited number of viral genes. The explanation lies, in part, in the utilization of host factors by viruses during their life cycle. Identifying these host factors that either facilitate or impede novel virus replication can uncover potential targets for antiviral therapies (Deol et al., 2022). A number of techniques, including forward genetic screens, are employed to study virus-host interactions. Although the CRISPR/Cas9 system is not the first tool for genetic screens, it has emerged as a highly robust method. Genome-wide CRISPR screens are being conducted to identify host genes involved in SARS-CoV-2 replication. For instance, Wang et al. identified ACE-2 as the entry receptor and highlighted TMEM106B, VAC14, cholesterol regulators, and exocyst subunits as additional host factors supporting SARS-CoV-2 infection, thus offering potential targets for antiviral strategies (Wang et al., 2021). Other researchers have also employed CRISPR-based screens to identify candidate host genes for SARS-CoV-2, paving the way for the development of CRISPR/Cas9-mediated antiviral therapeutics (Daniloski et al., 2021, Zhu et al., 2021). Huang et al. devised a CRISPR-based diagnostic approach that utilizes customized CRISPR Cas12a/gRNA complex and fluorescent probes to detect target amplifiers generated by
standard RT-PCR or isothermal recombinase polymerase amplification (RPA). This method enables sensitive detection at locations where the RT-PCR system necessary for qPCR diagnosis is unavailable (Huang et al., 2021). The technique exhibited high sensitivity, with a reaction time of approximately 50 minutes and a detection limit of two samples per sample.

The diagnostic results obtained using the CRISPR analysis from nasal swab samples of suspected COVID-19 patients were comparable to quantitative RT-PCR tests and outperformed standard clinical laboratory tests. With no specific treatment available for emerging infectious diseases like COVID-19, it is crucial to explore effective diagnostic and therapeutic approaches (Lou et al., 2022). A number of antiviral approaches focus on inhibiting different stages of the viral life cycle by targeting structural or non-structural components. CRISPR/Cas systems have gained significant attention as a means to limit viral replication by specifically targeting the viral genome. CRISPR/Cas9, in particular, has demonstrated its potential as an antiviral strategy against DNA viruses, both in vitro and in vivo (Lee, 2019). Similarly, the CRISPR/Cas12a system has shown promise in inactivating integrated HIV DNA genomes, surpassing Cas9 in HIV inhibition. On the contrary, the CRISPR/Cas13 system has emerged as a novel RNA-guided RNA targeting system, capable of recognizing and degrading the genomic RNA of SARS-CoV-2, thereby impeding virus replication. Recent research has demonstrated the efficacy of CRISPR-Cas13 in safeguarding host bacterial cells against phage infection through specific crRNA (Yan et al., 2019). This strategy can potentially be leveraged to design therapeutic drugs targeting the single-stranded RNA genomes of emerging infectious diseases. The Cas13 enzymes utilize a short hairpin crRNA to recognize specific sequences on the target RNA, and unlike Cas9, they do not require a protospacer adjacent motif (PAM) sequence (Burmistrz et al., 2020). Some subtypes of Cas13 also exhibit collateral cleavage activity, resulting in non-specific cleavage of target and non-target RNAs. This property has been utilized in diagnostic applications such as SHERLOCK. Different subtypes of Cas13 have been explored for their antiviral potential. Abbott et al. introduced a CRISPR-Cas13-based strategy known as PAC-MAN (Prophylactic Antiviral CRISPR in Human Cells), which effectively degrades SARS-CoV-2 RNA in cells by identifying functional crRNA specifically targeting SARS-CoV-2 (Lou et al., 2022). This method reduces viral load within cells, and a small set of six crRNAs can target over 90% of coronaviruses, making PAC-MAN a promising pan-coronavirus suppression strategy (Abbott et al., 2020). The development of CRISPR-based tools for the treatment of emerging infectious diseases holds great research prospects.

The primary drawback hindering the extensive use of the CRISPR-Cas system lies in its inability to accurately identify specific nucleic acids for diagnostic and therapeutic purposes (Doudna, 2020). To address this issue, one approach
The image illustrates the procedure of identifying SARS-CoV-2 using the CRISPR/Cas system. Initially, RNA is isolated from the individual using a conventional RNA extraction technique. In Cas12-dependent detection, the RNA is amplified into a duplex DNA, whereas in Cas13-dependent detection, the amplified DNA is transcribed into a solitary RNA strand. The collateral function of both Cas12 and Cas13 is employed for the detection of SARS-CoV-2. Image credits to Deol et al., 2022.

is to enhance the specificity of target nucleic acids by modifying the CRISPR protein, thus minimizing off-target effects (Naeem et al., 2020). Bioinformatics methods are commonly employed to detect off-target effects, and advancements in Cas9 have shown promise in reducing such effects (Lou et al., 2022, Coelho et al., 2020). Future research should focus on investigating off-target effects and refining detection techniques. In utilizing the CRISPR-Cas system for the prevention and control of infectious diseases, delivery tools play a crucial role in targeting cells within the body (Lou et al., 2022). However, the carrier function of CRISPR is limited by the size of viral genes, and most Cas proteins are relatively large in molecular weight (Liu et al., 2020). Overcoming this challenge necessitates the search for low molecular weight Cas proteins (Luther et al., 2018). Moreover, when Cas proteins derived from prokaryotes are administered to the human body, they can elicit an immune response and trigger the production of specific antibodies, which interfere with CRISPR’s immune response (Lou et al., 2022). Enhancing the efficiency of the CRISPR-Cas system is vital to mitigate the immune response in future CRISPR tool development. Cas9 and Cas12 proteins exhibit PAM-dependent recognition and cleavage of dsDNA, allowing them to target specific gene sequences within target genes (Karvelis et al., 2020).
However, some highly specific sequences may not be accessible, emphasizing the need for further enhancement of the CRISPR tool for target selection (Lou et al., 2022). In addition, the presence of RNA enzymes widely in existence leads to RNA instability, thereby affecting the diagnostic efficacy of CRISPR.

Conclusion

There has been considerable progress in molecular cloning techniques in research laboratories as well as widespread operation in medical microbiology. While DNA-based procedures have improved the diagnosis of diseases, conventional cell factories have drained their competence (Khan et al., 2016). It is obligatory to discover and integrate new production systems into the production process. In addition, a deeper understanding of the physiology of host cells and their responses to stress would enable enhanced tools by gene manipulation either at the genetic or at the metabolic levels for growing yield and eminence (Khan et al., 2016). The progressions in recombinant DNA technology, microbiology, genomics, bioinformatics, and related fields have, however, contributed significantly to the understanding of the pathogenic mechanisms underlying the microbial infectious diseases and how pathogens interrelate with hosts. Several new vaccine strategies have been established, as outcomes of these advances have produced promising results. Masses of children, globally, still die from infectious diseases notwithstanding the existence of currently available vaccines. Therefore, it is essential to understand the trials involved in developing recombinant vaccines and estimate the stability between charge, remunerations, and hazard before bringing a vaccine applicant to the hospital. As a result of this approach, patients with unrecognized or hard-to-diagnose infections will be identified and treated punctually, resulting in reduced stays and a reduction in iatrogenic events. Considering the relative costs of novel diagnostics compared to existing standards, public reimbursements will need to be carefully explored. Hence, molecular cloning can be delineated as a scientific effort to treat a lengthy list of diseases that have claimed a lot of human lives in the course of long period.

References


UMURJ • vol. 17, no. 1 • 2024


Unlocking the Power of Molecular Cloning: Revolutionizing Medical...


Drug treatment advancements for HIV have dramatically advanced since the virus’ identification in the early 1980s. Integrase strand transfer inhibitors (INSTIs) are one of the seven HIV treatment drug classes currently utilized to create an undetectable viral count in blood samples of people living with HIV (PLWH). First-generation INSTIs are documented with low barriers of genetic resistance, which indicates that the number of mutations to lead to a drug-resistant mutation is low. The introduction of dolutegravir, a second-generation INSTI, shows a higher barrier of genetic resistance that reduces drug-resistant mutations to INSTIs and increases the overall effectiveness of this class of HIV treatment. PLWH can be categorized based on whether they received treatment previously/currently or have never received treatment. Therapy-naive and previously treated (successfully or unsuccessfully) patients for HIV report different rates of drug-resistant mutations compared to actual resistance to dolutegravir, 0.4–31% and 0.1–67.2%, respectively. Evolutionary considerations of genetic resistance, including epistatic interactions and point mutations, suggest both non-polymorphic and polymorphic mutations for these drug-resistant mutations. An incomplete understanding of how evolutionary factors contribute to HIV drug resistance highlights the importance of conducting further research. This research may help improve the efficacy of second-generation INSTIs in future treatment options for PLWH. This review describes the landscape of existing research on drug resistance prevalence for dolutegravir and possible evolutionary explanations on how these mutations arise in the first place, leading to implications in developing more robust treatment modalities.
Background

The human immunodeficiency virus (HIV) is a prominent retrovirus identified in 1983 that has no cure, yet it can be readily treated with the right combination antiviral therapy (cART) (Greene, 2007). At the end of 2021, the World Health Organization (WHO) estimated that 38.4 million people around the globe are living with HIV (WHO, 2022). UNAIDS set Ambition 2030 targets to end the ongoing HIV epidemic: (1) 95% of people living with HIV (PLWH) know their status; (2) 95% of people who know their status are receiving treatment; (3) 95% of those receiving treatment are virally suppressed (Ehrenkranz et al., 2021). PLWH who are virally suppressed reach a level of virus in the body that cannot be detected by laboratory tests. At these levels, HIV cannot be transmitted to other individuals, highlighting relevant campaigns such as undetectable equals untransmittable (U=U) and treatment as prevention.

HIV is a very mutable virus (Yeo et al., 2020). Viral mutations in the body and during seroconversions to infection increase due to error-prone replication cycles. Therefore, the HIV virus can develop drug-resistant mutations, decreasing the effectiveness of these medications in high active antiretroviral therapy (HAART) regimens. The strain of HIV in PLWH can be naturally resistant to certain drugs due to certain mutations in the virus or due to the development of acquired resistance over time in PLWH. In addition, there is evidence to suggest that people who inconsistently adhere to their medication can see increased rates of drug-resistant strains of HIV (Chen, Chen, & Kalichman, 2017). When PLWH and HIV providers encounter drug resistance, the range of medications to treat an HIV diagnosis may become limited, impacting the health outcomes of PLWH.

Dolutegravir is a second-generation integrase strand transfer inhibitor (INSTI) approved by the US Federal Drug Administration (FDA) in 2013 to treat HIV (Kandel & Walmsley, 2015). As it entered the global state, many HIV providers and public health officials promoted its potential to be a more effective medication than existing first-generation INSTIs, which are well-tolerated, are easy to take, and have decreased drug-drug interactions (Rhee et al., 2019). Furthermore, the greatest factor for dolutegravir’s success is its higher barrier to genetic resistance compared to first-generation INSTIs like raltegravir, reducing the chance of an individual needing to switch treatment regimen throughout their lifetime. Even with dolutegravir’s increased barrier to resistance, there is currently little understanding of the population-level prevalence of dolutegravir drug resistance and how drug resistance arises.
Drug Resistance Prevalence

Drug resistance is broadly outlined as the reduction in effectiveness of medications in treating a disease or conditions with prime examples coming from the fields of antimicrobial resistance in antibiotics and cancer medicine (Habiboush & Guzman, 2022). PLWH can be categorized into two treatment categories: (1) therapy-naive patients and (2) previously treated patients (Colombo, DiMatteo, & Maggiolo, 2013; Tseng, Seet, & Phillips, 2015). Both therapy-naive and previously treated patients are individuals who have seroconverted, meaning the body has responded to HIV by creating antibodies. Importantly, therapy-naive patients contain a level of the virus in the body that is detectable for treatment response yet have never received treatment for HIV. Previously treated patients for HIV are individuals enrolled in previously successful or unsuccessful treatment regimens.

Research suggests that the rate of drug resistance for INSTIs in therapy-naive patients is 3.82%, while HIV in previously treated patients is resistant at 11% (Fan et al., 2022; Kamelian et al., 2019). For therapy-naive patients, the genotypes of the individual’s virus were screened for resistance-associated mutations (RAMs), which may suggest an individual’s inherent resistance to INSTIs. RAMs can be major or accessory. Major RAMs (Y143R/C/D/G and P145S) are shown to create actual drug resistance. Accessory RAMs (G140E, E157Q, and G163R) can, in combination and over time, lead to actual drug resistance. Among samples of therapy-naive patients in Cameroon, Mikaski et al. reported 5.4% major RAMs and 8.1% accessory RAMs (Mikasi et al., 2020). It is important to note that these samples were conducted for all INSTI mutations, not dolutegravir specifically. Áy et al. (2021) collected data on therapy-naive patients, which report different percentages of mutations, including 1 out of 249 (0.4%) to have major RAMs and 31% accessory RAMs. These dramatically different percentages may reflect population or geographically distinct prevalence of major and accessory RAMs associated with dolutegravir. These samples identify the critical role drug resistance screening plays in the larger rollout of dolutegravir and emphasize the research that must be conducted to pre-screening efforts.

There is markedly more research for drug resistance among previously treated patients. First-generation INSTIs have shown a plethora of drug resistance mutations leading to the increased failure of treatment regimens for HIV (Anstett et al., 2017). Data collected for drug resistance to dolutegravir are separated into two main outcomes for PLWH: (1) potential resistance from possible future mutations that can reduce the effectiveness of dolutegravir and (2) inherent resistance from existing mutations to dolutegravir in a specific HIV strain. Importantly, previously treated patients can be categorized based on their success or failure of previous treatment regimens. The range of actual resistance in PLWH was from 0.1–0.7% to 21.9% among those who are currently successfully virally suppressed under
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamelian et al., 2019</td>
<td>HAART-treated individuals</td>
<td>British Columbia, Canada</td>
<td>0.1–0.7% (1 to 7 per 1000)</td>
<td>–</td>
</tr>
<tr>
<td>Van Oosterhout et al., 2022</td>
<td>First-generation INSTI regimen failure</td>
<td>Malawi</td>
<td>29.6% (8 out of 27)</td>
<td>–</td>
</tr>
<tr>
<td>Fourati et al., 2015</td>
<td>First-generation INSTI regimen failure</td>
<td>France</td>
<td>13.9%</td>
<td>64%</td>
</tr>
<tr>
<td>Engone-Ondo et al., 2021</td>
<td>HAART-treated individuals; First-generation INSTI regimen failure</td>
<td>Semi-rural Gabon</td>
<td>21.9%; 67.2%</td>
<td>84.6% (among patients who failed first-generation INSTI regimen)</td>
</tr>
</tbody>
</table>

Table 1: Description of Actual and Potential Resistance from Current Literature Categorized by Patient and Geographic Populations

HAART (Kamelian et al., 2019; Engone-Ondo et al., 2021). These percentages greatly increased for PLWH who had previously or are currently failing first-generation INSTI treatment regimens, ranging from 29.6% to 67.2% showing resistance to dolutegravir (van Oosterhout et al., 2022; Engone-Ondo et al., 2021). Critically, 64–84.6% of HIV sequences screened show potential drug resistance mutations to dolutegravir among those who previously or are currently failing first-generation INSTIs (Fourati et al., 2015; Engone-Ondo et al., 2021). Saladini et al. (2012) reported that 59.8% of samples collected from previously treated patients harbored at least one of the resistance mutations for first- or second-generation INSTIs. Table 1 consolidates the patient and geographic population with the actual and potential resistance described in the source paper. Nonetheless, this does not indicate actual resistance, nor does it show treatment regimen failure for dolutegravir for PLWH.

**Evolutionary Explanation to Drug Resistance**

**HIV Drug Evasion**

Drug resistance among HIV strains in PLWH can vary in their categorizations, including transmitted, acquired, or multi-class drug resistance (Pennings, 2013).
Transmitted drug resistance indicates a viral strain that, when transmitted to another host, already contains drug-resistant mutations, while acquired drug resistance explains how an individual’s strain of HIV can mutate over time to confer drug-resistant mutations. Multi-class drug resistance describes strains of HIV that confer multiple mutations that evade more than one of the seven drug classes that target HIV. There is a stark geographic and socioeconomic influence of transmitted drug resistance, where the standard for patients in higher-income countries is to screen for genetic resistance before treatment begins. By conducting genetic screening, a provider can readily assess the feasibility of PLWH to receive some treatment over others. Acquired drug resistance appears to increase over time for a treated patient, indicating that although not all patients over time will develop resistance, a small subset of the population will develop drug resistance HIV virus strains. This type of resistance is notable for potential dolutegravir candidates who may have previously failed a first-generation INSTI in the same drug class. cART requires drugs from multiple classes, and when PLWH are limited in access to multiple drug classes, the course of treatment over a lifetime may become limited, creating future problems for maintaining viral suppression. Research suggests that drugs targeting viruses like HIV may have imperfect tissue penetrations and result in spatial monotherapy (Moreno-Gomez et al., 2015). These implications may stand to explain how some HIV strains in individuals adapt in environments where medication is not present.

Genetic exchange, re-assortment, and recombination of HIV could contribute to its adaptation at the population genetics level (Wilson et al., 2015). While the HIV virus on its own can mutate to adapt, the question of human adaptation to HIV drug resistance is largely unstudied, including the possible influences of linkage disequilibrium and epistatic interactions. Hence, there is an emphasis to assess standing genetic variation in new treatment initiation for PLWH to determine whether drug-resistant mutations are present.

First-Generation INSTIs

First-generation INSTIs are shown to have a low genetic barrier to resistance with the medications raltegravir and elvitegravir (Anstett et al., 2017). The field of virology and resistance in HIV have distinctly categorized the type of mutation from conventional evolutionary terminology. Non-polymorphic mutations are defined by a percentage of mutations occurring less than 1% of any subtype of HIV virus (Rhee, Tzou, & Shafer, 2021). Polymorphic mutations are sites with variable frequency by which the sequence of a gene is found in more than 1% of the population. The occurrence of drug resistance mutations among first-generation INSTIs appears to occur through the transmission of drug mutations rather than
naturally arising resistance mutations in PLWH. Áy et al. (2021) characterize these drug-resistant mutations as non-polymorphic. In addition, Saladini et al. (2012) provide further evidence that first-generation INSTI naive patients did not have mutant strains that contained drug resistance mutations as natural polymorphisms. Two mutations (T124A & L101I) were detected among naive and treated patients, yet their prevalence was the same for either group. However, other research in Chinese populations has found polymorphic accessory mutations, which could cause low-level resistance to first-generation INSTIs (Yu et al., 2022).

**Second-Generation INSTIs**

The resistance profile of dolutegravir is extensively characterized by Rhee et al. (2019). The development of resistance mutations to dolutegravir has been observed in vivo and in samples from the populations (Fourati et al., 2015). Different mutations are categorized as directly impacting the effectiveness of dolutegravir and creating possible pathways that could lead to second-generation INSTI drug-resistant mutations.

Anstett et al. (2017) suggest that the increased resistance profile of dolutegravir is associated with a longer binding half-life, which maintains activity against more resistant first-generation strains. This further supports a theory of different binding properties to explain dolutegravir’s resistance profile (Garrido et al., 2011). However, the authors do not present an evolutionary mechanism for why this

<table>
<thead>
<tr>
<th>Source</th>
<th>Mutation</th>
<th>Mutation Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anstett et al., 2017</td>
<td>N155; Q148</td>
<td>Resistant; Pathways to Resistance</td>
</tr>
<tr>
<td>Pham et al., 2021</td>
<td>S153F or S153Y with R263K</td>
<td>Resistant</td>
</tr>
<tr>
<td>Garrido et al., 2011</td>
<td>T124A and L101I+T124A</td>
<td>Resistant</td>
</tr>
<tr>
<td>Brenner et al., 2016</td>
<td>G118R</td>
<td>Resistant</td>
</tr>
<tr>
<td>Saladini et al., 2012</td>
<td>L101I and T124A</td>
<td>Resistant</td>
</tr>
<tr>
<td>Rhee et al., 2019</td>
<td>R263K, G118R, N155H, Q148H/R; Q148 + G140 and/or E138</td>
<td>Resistant</td>
</tr>
<tr>
<td>George et al., 2018</td>
<td>T97A</td>
<td>Resistant</td>
</tr>
<tr>
<td>Ndashimuye et al., 2018</td>
<td>E157Q</td>
<td>Pathways to Resistance</td>
</tr>
</tbody>
</table>

**Table 2**: Description of Mutation and Mutation Types from the Current Literature
type of advantage is displayed by dolutegravir. Pham et al. (2021) offer that experiments conducted in their study show that deleterious effects of individual substitutions in the integrase codon region could lead to possible dolutegravir drug resistance. The studies listed in Table 2 largely characterize singular substitutions as a leading cause to genetic resistance, but some studies indicate that the epistasis can occur alongside these drug-resistant mutations. Epistasis occurs when the effect of a gene mutation is dependent on the presence or absence of mutations in or more other genes (Churchill, 2013; Garrido et al., 2011). Both Garrido et al. (2011) and Brenner et al. (2016) suggest that natural polymorphism in different HIV strain subtypes is partially responsible for drug resistance to dolutegravir for PLWH.

Conclusion

There is some prevalence in therapy-naive and previously treated patients of drug-resistant strains and mutations to the medication dolutegravir for PLWH. Most concerning is the high percentage of dolutegravir resistance among individuals currently taking or previously failing first-generation INSTIs, suggesting that alternative HIV medication routes may be necessary for these PLWH. Although resistance is prevalent, there is little research on the evolutionary mechanisms that may cause these drug-resistant mutations or strains. Broadly, first-generation INSTIs confer largely non-polymorphic mutations in creating dolutegravir drug resistance, while natural polymorphism is used as a primary explanation for the presence of drug-resistant mutations in second-generation INSTIs. Second-generation INSTI drug-resistant mutations appear to arise from individual substitutions with possible epistatic associations among some mutation locations.

These drug-resistant mutations against dolutegravir are critical in ending the HIV epidemic. INSTIs are quickly leading the charge on long-acting medications, while biomedical advancements in other HIV drug classes are lagging behind. Cabotegravir is a second-generation INSTI that is derived from dolutegravir and acts as a long-acting injectable medication versus a daily oral pill (Landovitz et al., 2021). If individuals do not have access to long-acting medications due to drug-resistant mutations limiting their drug class options, the treatment for PLWH may become challenging. Without further characterization and screening of drug-resistant mutations for second-generation INSTIs, potential biomedical advancements may face drug-resistant mutations in existing strains of HIV circulating in PLWH. Therefore, more research is needed to determine the evolutionary origins of drug-resistant mutations to INSTIs.
References


World Health Organization. (2022, July). HIV. The Global Health Observatory. https://www.who.int/data/gho/data/themes/hiv-aids#:~:text=Globally%2C%2038.4%20million%20in%2033.9%20%5B33.9%E2%80%93,considerably%20between%20countries%20and%20regions.https://www.who.int/data/gho/data/themes/hiv-aids#:~:text=Global%2C%2038.4%20million%20in%2033.9%20%5B33.9%E2%80%93,considerably%20between%20countries%20and%20regions.


A Study on the Effect of Zamzam Water on the Germination and Growth of Jasminum sambac

Bilal Irfan and Ihsaan Yasin

This paper seeks to investigate the role of Zamzam water in enhancing the germination and post-germination growth of Jasminum sambac. Zamzam water contains many mineral and chemical properties that have been previously tested as a model for supporting fertilization efforts in agricultural production. Seeds were planted for Zamzam water with a distilled water serving as a control to tabulate the rate of germination and the length of emergent seedlings. Zamzam’s role in supporting the growth of Jasminum sambac was displayed with statistically significant differences being present, suggesting potential for further research into Zamzam’s potential as an agent for plant growth.

Introduction

Jasminum sambac, commonly known as the Arabian jasmine, but also known as sampaguita or melati putih, is a small shrub or vine that is native to tropical Asia and is found in a number of desert-filled regions experiencing a shortage of water. Saudi Arabia is one of the many countries in the continent that cultivate Jasminum sambac for its enchanting and fragrant flowers. The country is known for experiencing large seasons of drought or periods with limited rainfall, which could affect the usage of the Zamzam Well in Mecca as a supplemental water source for irrigation practices. The well is located at a depth of 30 meters within Masjîd Al-Haram, the holiest site for Muslims, and is believed to have been found by Hajar while caretaking for her son Ismail during the time of Abraham in the Islamic tradition. This experiment, centered around biological processes of plants, aimed to study the various effects of Zamzam water on the germination and growth of Jasminum sambac seeds.

Contact: Bilal Irfan <birfan@umich.edu>

https://doi.org/10.3998/umurj.5511
Past studies have shown positive results of Zamzam water when used in the production of food products, including broad beans and wheat (Mutwally et al., 2015; Mohammad & Al Hatem, 2022; Elmoula et al., 2020). Using Zamzam water as a stand-alone, or in congruence with distilled water, resulted in significantly observed increases in the percentage of seed germination, shoot length, and the fresh and dry weights of the shoots. On that note, in contrast to broad bean plants irrigated by other water sources, the ones using Zamzam water had a noticeably higher display of flowers. Research by Hamed et. al. (2009) into the yields of Vicia faba L. and Triticum vulgare L. found that harvests inundated with Zamzam water gave the most elevated estimations of yield, dissolvable starches, complete sugars, protein, and all-out nitrogen substance. Furthermore, Alsokari found a beneficial outcome on the development and protein substance of lentil seedlings when watered with that out of the Zamzam Well (Alsokari, 2011). The water system with Zamzam water expanded the protein, RNA, DNA, and the phenolic and antioxidant agent substances in lentil seedlings (Mardi et al., 2015). The primary inquiry motivating this study revolved around investigating whether Zamzam water elicits a notable disparity in the germination and subsequent growth of Jasminum sambac seeds when compared to distilled water.

Our prediction suggests that Zamzam water will significantly enhance the germination of Jasminum sambac compared to distilled water. This can be attributed to the abundance of chemicals such as calcium, magnesium, sodium, and chloride in Zamzam water, which are instrumental in enhancing predictions regarding its differentiations from distilled water (Donia & Mortada, 2021). These minerals serve as the paramount divalent cations that are essential to reacting with the multifaceted carbonates and bicarbonates in seeds and essential to the following processes under germination: enzyme activation, nutrient transport, and protein synthesis. Therefore, we can anticipate that germination. Hence, we can expect germination to be enhanced using Zamzam water, and specifically, this can be viewed by the root length should germination occur for a given sample.

Methods and Materials

Preliminary Experiment

To ensure this experiment was viable, and to observe firsthand the beginning processes of germination, Jasminum sambac seeds were purchased, while Zamzam water was transported from Mecca, Saudi Arabia. Twenty seeds were planted across two petri dishes, with a diameter of 5 cm each, of Miracle-Gro soil (ten seeds per dish). Each dish was given 15 ml of distilled water and left for
one week at room temperature (range of 60°–75°F) before being watered again (for a total of two weeks). Following the conclusion of the two-week time frame, the number of germinating seeds was counted. Twelve of the seeds germinated for a total of 60% in germination results. This showed that the seeds do, in fact, germinate and would be viable for this experiment to use. The aforementioned protocol was organically developed to provide a standardized, replicable environment that would allow for accurate observation and measurement of the germination processes of *Jasminum sambac* seeds, thereby ensuring the consistency and validity of the results.

**Experimental Setup and Germination**

Ten petri dishes, 150 g of Espoma organic potting soil, 100 seeds of *Jasminum sambac* that are 2.50 mm in diameter (±0.5mm), 100 ml of Zamzam water, 100 ml of distilled water to create the control, two sand baths, and two thermometers (±0.05°C) were required for an initial setup. Sand baths serve to create a level surface for planting and to provide consistent moisture to the seeds. The sandbags help to keep the soil in place and prevent erosion while also retaining water and promoting seed germination within a stable and controlled environment. The sand baths were set at 78°F and a thermometer placed in each one to authenticate the desired temperature setting. Five petri dishes were placed into one sand bath to be tested, with the remaining five dishes in another to be utilized as a control. Measurements of 15 × 10.0 g of the soil, while making use of the electronic weighing scale, and placement of 10.0 g into each one of ten petri dishes occurred. Five of the dishes were reserved to be tested with Zamzam water, while the remaining five dishes were treated with distilled water as the control. Then, ten *Jasminum sacrum* seeds were planted into each one of the petri dishes and planted into the prescribed soil at a consistent depth of 0.8 cm. Each of the respective sand baths was watered at a consistent time in the evening (6:00 p.m.) with 15 ml of either distilled or Zamzam water every 3 days for 15 days. After 15 days, the number of seeds that germinated was counted (noted from the emanation of the seedling) and measured to assess the height of the emergent seedling in the test and the control groups with the 10.0 cm ruler. The seedling height was taken from the soil surface to the highest visible part of the stem. The process outlined here was repeated to ensure that sufficient data was present for this experiment.

To ensure that the controlled variables served as a placeholder for tabulating precise germination data, the same amount of water, 15 ml of volume, was added to each petri dish simultaneously at the designated evening time every 3 days throughout the 15-day period. All 150 *Jasminum sambac* seeds used in this
experiment were kept within a size range of 2.50 mm in diameter. The temperature of the seeds was kept constant at 78.0°F by the sand baths and was checked via thermometers every day to ensure consistency. All seeds were planted in Espoma organic potting soil. With the soil held constant, it is reasonable to infer that its contents contain similar concentrations of its various nutrients. The mass of the soil was also kept constant at 15.0 g. Near-same amounts of light were presumed to be received for each seed as the entirety of the experiment was conducted in the same location in front of a window. In addition to these precautions taken for ambient light, the seeds were consistently planted at a depth of 0.8 cm and near the edge of the petri dish, enabling an easy observation of its changes through the presence of the glass in lieu of digging the seeds up every time.

Some of the underlying assumptions going into the experiment included that the penetration/input of light through the petri dishes is the same intensity due to the physical proximity of the sand baths, the soil has the same concentration of its various nutrients and elements, any impurities and chemicals found in the air particles of the room will be the same for the two sand baths, and the *Jasmine sambac* seeds contain the same percentage of their respective elements.

The seeds that germinated from each of the petri dishes were counted from among the test and control groups by observing from the side of the glass if the seedling’s outer coat had broken and if the plant’s initial parts had emerged. Measurements of the height of the seedling (determined by the distance from the tip of the seedling to the soil surface) of the seeds that are germinated after a 15-day period were also taken. The $\chi^2$ test was used to assess differences in germination, and the t-test was used to assess differences in the growth of the seedlings.

### Germination Results

#### Observations

By the end of the experiment, the *Jasminum sambac* seeds were about 1–1.5 cm (100–150 mm) in length, with a width of about 0.5 cm in the middle. The seeds had an ovoid shape with a little notch in a circular shape on the top of them. The Zamzam water and distilled water were distinct in their effect on the seeds. The Zamzam water had a mild odor like the smell of the ocean from a distance. There was significantly more germination in the seeds that were planted with Zamzam water than those planted with distilled water. While there was less than expected uniformity in germination time between
the *Jasminum sambac* seeds planted in Zamzam or distilled water, it was noted that those planted with Zamzam showed signs of seed cracking and emergence of root within four to six days, compared to around eight days for those with distilled water. Seeds watered with distilled water remained a licorice color for most of the time, only shifting to a maroon brown in some cases, while those left in Zamzam water showed slight lightening of the color to a light gray mixture in the black, as if the color of the seed was being drained out. Though only small roots emerged at any point from the seeds with a little white coating, germination developed further with Zamzam. There was a larger opening and larger roots sprouting in the seeds planted with Zamzam water. No leaves ever emanated in the two-week time frame from either the test or the control, indicating that they may have needed to grow longer.

**Number of Successfully Germinated Seeds**

To determine how many seeds could be classified as successfully germinating, a thorough analysis of both the cracking of the seed coat and the emergence of a root quantified germination for the purposes of this experiment. Those results have been recorded in Table 1.

Thirty-four of 50 seeds planted in Zamzam water germinated, a 68% success rate and an average of seven seeds germinating in each petri dish. Twenty-two of 50 seeds planted in distilled water germinated, a 44% success rate with an average of four seeds germinating in each petri dish. It is worth noting that this 44% rate fell below the 60% recorded in the pre-experimental trials involving distilled water. See Table 2.

<table>
<thead>
<tr>
<th>Trial # (individual petri dishes)</th>
<th>Seeds Germinated with Zamzam Water (out of 10)</th>
<th>Seeds Germinated with Distilled Water (out of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

*Table 1: Germination Statistics*
Chi-Square Test

A chi-square test was conducted to see if there was a significant difference between the germination growth of seeds planted with Zamzam versus distilled water. The null hypothesis stipulated that “Zamzam water does not affect the germination of *Jasmine sambac* seeds.” The alternative hypothesis was “Zamzam water does affect the germination of *Jasmine sambac* seeds.” See Table 3.

While values of the final column were rounded to three significant figures in the table, original values were used for the total on the bottom of 5.84 (also thereby rounded to three significant figures for the purpose of this report).

To calculate the number of degrees of freedom = (rows − 1) × (columns − 1), which in this case would translate to $(2 - 1) \times (2 - 1) = 1$

$$x^2_{crit} = 3.84 \text{ at } p = 0.05$$

Due to $x^2_{calc} = 5.84$ being larger than the $x^2_{crit} = 3.84$, the null hypothesis must be rejected in favor of the alternative hypothesis. The test value adopted was significant at some point between $0.01 < p < 0.05$. It would also further remain significant at $p < 0.01$. From these results, we were able to conclude that Zamzam water does affect the rate of germination in *Jasmine sambac* seeds.

Post-germination Growth of the *Jasmine sambac*

Having determined that Zamzam water does affect the rate of germination, a further test was conducted to measure post-germination growth. For this follow-up test, both distilled and Zamzam water were utilized to investigate the effectiveness of Zamzam water. The parameters of effectiveness were defined to be the length of the roots growing out of a seedling from successfully germinated seeds. Any additional growth from the seedling was construed as more effective for the purposes of this experiment. The initial raw data used to assess the processed results later can be found in the appendix. See Table 4 and Figure 1.
Table 3: Chi-Square Test

<table>
<thead>
<tr>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
<th>Difference</th>
<th>Absolute Difference</th>
<th>$(O-E)^2$</th>
<th>$(O-E)^2/E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>E</td>
<td>O−E</td>
<td></td>
<td>O−E</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>28</td>
<td>−6</td>
<td>6</td>
<td>36</td>
<td>1.29</td>
</tr>
<tr>
<td>22</td>
<td>28</td>
<td>6</td>
<td>6</td>
<td>36</td>
<td>1.29</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>6</td>
<td>6</td>
<td>36</td>
<td>1.64</td>
</tr>
<tr>
<td>28</td>
<td>22</td>
<td>−6</td>
<td>6</td>
<td>36</td>
<td>1.64</td>
</tr>
</tbody>
</table>

$X^2$ calculation 5.84

Table 3: Chi-Square Test

Table 4: Height of Seedlings for Germinated Seeds

<table>
<thead>
<tr>
<th>Water Type</th>
<th>Average Height in mm (±0.5 mm)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zamzam</td>
<td>9.65</td>
<td>3.58</td>
</tr>
<tr>
<td>Distilled</td>
<td>6.5</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 4: Height of Seedlings for Germinated Seeds

328 is the combined height of all the seeds that germinated with Zamzam water, and 143 is the combined height of all the seeds that germinated with distilled water.

Figure 1: Average Seedling Length vs. Water Type
T-test: A two-tailed test was conducted to determine whether the seeds planted in Zamzam water germinated more (as in more root material came out) than the seeds planted in distilled water.

The null hypothesis proposed “Zamzam water has no effect on the post-germination growth of *Jasmine sambac* seeds.” The alternative hypothesis postulated that “Zamzam water does have an effect on the post-germination growth of *Jasmine sambac* seeds.”

Discussion

*Over two weeks, this experiment yielded a 68% germination rate of* *Jasmine sambac* seeds *planted in Zamzam water compared to the 44% rate of distilled water.* A further analysis of these results reveals that the average height (in mm) for seeds placed in Zamzam water after two weeks was greater than those put in distilled water. Furthermore, there is a higher standard deviation among those placed in Zamzam water, which could be attributed to a variety of reasons. This difference was also present in the raw data tables, suggesting that there may be some level of significant difference between the post-germination growth of seeds placed in Zamzam versus distilled water. Therefore, a t-test was conducted to ascertain the magnitude of any such difference.

The two-part hypothesis stating that Zamzam water would aid in enhancing germination and post-germination for the *Jasminium sambac* in comparison to distilled water holds true from the results obtained, and there is a significant relationship between Zamzam and the effectiveness of germination. The test value of 5.03 is greater than the $t_{crit} = 2.872$ at $p = 0.05$, indicating a likely relationship between Zamzam water and its ability to project and stimulate growth in *Jasmine sambac* seeds. The null hypothesis can be rejected.

Given the positive growth of the seeds under Zamzam water, this difference is highly favorable toward Zamzam water. A further analysis will need to be conducted to determine the true efficacy of Zamzam over distilled water. Regardless, our prediction and subsequent reasoning are supported by the results of the experiment.

While much of the experiment was done with adequate controls, there were certain areas wherein changes could be made to allow for the experiment to be done in a more concise and effective manner. Espoma organic potting soil was utilized in this experiment for both the seeds that were planted with Zamzam and distilled water. However, this soil is enhanced with myco-tone, which contains elements of endomycorrhizae and ectomycorrhizal acting as a fertilizer to promote root growth, specifically, which was analyzed for post-germination in this study. Zamzam water itself contains several chemicals including dissolved
salt, sodium, calcium, and magnesium. Its pH level differs from that of distilled water, ranging from 6.5 to 8.5 versus 7 (Donia & Mortada, 2021). These properties may have an impact on the soil or the water, resulting in an imperfect attempt to ascertain what was truly caused by Zamzam water or not. Further study is warranted to examine differing conditions.

Future experiments can mitigate these variations by testing multiple soil samples with differing chemical compositions, including flat clean soil or mulch, tested in the same two-week time frame of the experiment to measure growth. Some uncertainty is expected with easily accessible outside soil having a variety of impurities unaccounted for present within them and would need to be measured in a lab for maximum accuracy.

This experiment could also be enhanced by utilizing different types of Zamzam water. Companies that sell mixtures of Zamzam and distilled water, to avoid excess use of pure Zamzam, could be scouted out for purchases. A third trial set of 50 seeds, containing a mixture of equal parts of Zamzam and distilled water, could produce results displaying a clearer picture of how much Zamzam affects the rate of germination. This trial could also have iterations of the parts of Zamzam or distilled water to measure the impact of Zamzam’s properties on germination and post-germination to measure the maximum impact ratio. In addition, water with differing pH levels or with different chemical composition could be utilized in separate trials to determine which properties of Zamzam are enhancing growth.

While the seed size was mostly uniform length, another variation of testing could be seeds measured by weight, as there was some natural variation in the sample size. Seed weight was not factored into this experiment, and further studies could explore potential effects on germination and post-germination growth. Adoption of a universal control element, with precise weights and lengths, such as 1 cm, could be used to reduce possible fluctuations in results.

Difficulties in ensuring seeds were uniformly planted at a depth of 0.8 cm were present due to the surface layer of the soil not being perfectly evenly distributed and flat. Whether this impacted the trial significantly is uncertain, but further testing with a larger sample size could lead to more effective results. The 15-day time frame of this experiment and the limitation of testing only *Jasmine sambac* are restraining factors in a more cohesive understanding of the effects of Zamzam water. Additional investigations examining germination trends for Zamzam water across different specimens can be formulated to map out models for Zamzam’s role as an agent of enhancing plant growth.

The nutrient-rich Zamzam water, as well as the interrelations of its various elements, stimulates plant growth. The naturally occurring salt content within Zamzam water could influence the growth observed in post-germination due to how it may have reacted with the soil. A study by Algandaby M. Mardi and Al-Zahrani S. Hassan (2015), working for the Department of Biological Sciences
in the Kingdom of Saudi Arabia, found that varying concentrations of Zamzam water can universally help germination in many different plants, for irrigation or other purposes. They found that, using Zamzam water, Sesamum indicum as a field crop grew at a faster rate than with distilled water or other local water sources. In addition, they found higher protein content in plants given Zamzam water. Potassium chloride, known for germination and the seed-cracking process of plants, is one of the salts found in Zamzam.

The effects of Zamzam are of interest to researchers, who have conducted a spectrum of experiments ranging from anticancer properties and antimicrobial activity to mice offspring production. One study attempting to ascertain the role of potential pesticides found that when compared to globe artichoke extract alone, Zamzam water enhanced the characteristics of the extract and increased death of P. solenopsis, while Zamzam water alone induced mortality and was effective as a stand-alone pesticide (Abd-Allah, 2022).

Conclusion

There are clearly substantial effects of Zamzam water on plant germination and growth as well as agriculture. The water’s high mineral concentration makes it a fantastic source of nutrients for plants, encouraging quicker germination, longer shoots and roots, and more biomass buildup. This might have an impact on food security and sustainability since it could offer a different supply of irrigation water for crops, particularly in areas with a scarcity of freshwater resources. In addition, Zamzam water’s antimicrobial qualities may reduce the need for pesticides and other chemical treatments to control plant diseases, enhancing food safety and lowering agriculture’s environmental impact.

However, Zamzam water’s creation through synthetic means and potential for commercialization should be handled delicately. The well has important cultural and religious connotations and is a holy place for Muslims. The well is also a significant historical monument for the Hejaz region and is said to be around 4,000 years old. Any commercial use of the water would have to consider how crucial it is to protect the well and its surroundings for future generations. Overall, even though Zamzam water has a lot of potential for agricultural use, any commercialization efforts should carefully consider its special qualities and cultural significance.

Acknowledgments

We thank the following individuals for their substantial contributions in serving as advisors and critically reviewing the study proposal: Amanda McSween, MA,
and Jaclyn Toepel-Hartzell, MA. We would also like to thank Jumanah Saadeh, MA, for her role in the technical editing of the manuscript.

References


# Appendix

<table>
<thead>
<tr>
<th>Seed Number</th>
<th>Did the Seed Germinate</th>
<th>Height of Seedling in/mm ±0.5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>13.0</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>11.0</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>10.0</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>10.0</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>12.0</td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>9.0</td>
</tr>
<tr>
<td>11</td>
<td>Yes</td>
<td>8.0</td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>13</td>
<td>Yes</td>
<td>13.0</td>
</tr>
<tr>
<td>14</td>
<td>Yes</td>
<td>14.0</td>
</tr>
<tr>
<td>15</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>16</td>
<td>Yes</td>
<td>3.0</td>
</tr>
<tr>
<td>17</td>
<td>Yes</td>
<td>5.0</td>
</tr>
<tr>
<td>18</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>19</td>
<td>Yes</td>
<td>2.0</td>
</tr>
<tr>
<td>20</td>
<td>Yes</td>
<td>4.0</td>
</tr>
<tr>
<td>21</td>
<td>Yes</td>
<td>11.0</td>
</tr>
<tr>
<td>22</td>
<td>Yes</td>
<td>11.0</td>
</tr>
<tr>
<td>23</td>
<td>No</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 1: Results of Seeds Watered with Zamzam Water
<table>
<thead>
<tr>
<th>Seed Number</th>
<th>Did the Seed Germinate</th>
<th>Height of Seedling in/mm ±0.5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>25</td>
<td>Yes</td>
<td>8.0</td>
</tr>
<tr>
<td>26</td>
<td>Yes</td>
<td>8.0</td>
</tr>
<tr>
<td>27</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>28</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>29</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>30</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>31</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>32</td>
<td>Yes</td>
<td>15.0</td>
</tr>
<tr>
<td>33</td>
<td>Yes</td>
<td>17.0</td>
</tr>
<tr>
<td>34</td>
<td>Yes</td>
<td>14.0</td>
</tr>
<tr>
<td>35</td>
<td>Yes</td>
<td>9.0</td>
</tr>
<tr>
<td>36</td>
<td>Yes</td>
<td>13.0</td>
</tr>
<tr>
<td>37</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>38</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>39</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>40</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>41</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>42</td>
<td>Yes</td>
<td>11.0</td>
</tr>
<tr>
<td>43</td>
<td>Yes</td>
<td>10.0</td>
</tr>
<tr>
<td>44</td>
<td>Yes</td>
<td>9.0</td>
</tr>
<tr>
<td>45</td>
<td>Yes</td>
<td>15.0</td>
</tr>
<tr>
<td>46</td>
<td>Yes</td>
<td>12.0</td>
</tr>
<tr>
<td>47</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>48</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>49</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>50</td>
<td>Yes</td>
<td>11.0</td>
</tr>
</tbody>
</table>

*Table 1: (Continued)*
<table>
<thead>
<tr>
<th>Seed Number</th>
<th>Did the Seed Germinate</th>
<th>Height of Seedling in/mm ±0.5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>9.0</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>8.0</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>13</td>
<td>Yes</td>
<td>10.0</td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>16</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>17</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>18</td>
<td>Yes</td>
<td>9.0</td>
</tr>
<tr>
<td>19</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>20</td>
<td>Yes</td>
<td>5.0</td>
</tr>
<tr>
<td>21</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>22</td>
<td>Yes</td>
<td>4.0</td>
</tr>
<tr>
<td>23</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>24</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>25</td>
<td>Yes</td>
<td>3.0</td>
</tr>
<tr>
<td>26</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>27</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>28</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>29</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>30</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>31</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>32</td>
<td>Yes</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Table 2: Results of Seeds Watered with Distilled Water
<table>
<thead>
<tr>
<th>Seed Number</th>
<th>Did the Seed Germinate</th>
<th>Height of Seedling in/mm ±0.5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>34</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>35</td>
<td>Yes</td>
<td>10.0</td>
</tr>
<tr>
<td>36</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>37</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>38</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>39</td>
<td>Yes</td>
<td>3.0</td>
</tr>
<tr>
<td>40</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>41</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>42</td>
<td>Yes</td>
<td>6.0</td>
</tr>
<tr>
<td>43</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>44</td>
<td>Yes</td>
<td>7.0</td>
</tr>
<tr>
<td>45</td>
<td>Yes</td>
<td>4.0</td>
</tr>
<tr>
<td>46</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>47</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>48</td>
<td>Yes</td>
<td>1.0</td>
</tr>
<tr>
<td>49</td>
<td>No</td>
<td>0.0</td>
</tr>
<tr>
<td>50</td>
<td>Yes</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Table 2: (Continued)

Calculations for % Germination of Seeds

56% of all seeds germinated:

\[
56 / 100 = 0.56, \text{ multiplied by } 100 \text{ for a } 56\% \text{ overall germination rate. } 100\% - 56\% = 44\%; \text{ 44\% of seeds did not germinate.}
\]

The number of seeds planted in Zamzam water that were expected to germinate was calculated as: 56\% of 50 = 28.

The number of seeds planted in distilled water that were expected to germinate was calculated as: 56\% of 50 = 28.

The number of seeds planted in Zamzam water that were expected to not germinate was calculated as: 44\% of 50 = 22.

The number of seeds planted in distilled water that were expected to not germinate would be calculated by: 44\% of 50 = 22.
Association between Coexisting Hypertension, Dyslipidemia, and Elevated C-Reactive Protein with Cardiovascular Disease in the Health and Retirement Study

Thomas Karadimas and Helen C.S. Meier

Background

Cardiovascular disease (CVD) is the leading cause of death globally. Hypertension and dyslipidemia are established CVD risk factors, but these diagnoses are often insufficient in predicting CVD individually. Inflammation has emerged as a contributor to CVD, but research on the combination of inflammation, hypertension, and dyslipidemia in CVD risk is limited. More thorough evaluations of CVD risk using all 3 aforementioned risk factors are crucial.

Methods

This report analyzes data from the Health and Retirement Study, a representative cohort of US adults over 50 years of age (n = 1,527). Participants were classified as having prevalent CVD if they self-reported a healthcare provider’s diagnosis of either a heart condition or a stroke or transient ischemic attack (TIA). We developed a CVD risk score using three factors known to contribute to CVD: hypertension, dyslipidemia, and elevated C-reactive protein (CRP). Risk was categorized as low (0–1 factors), medium (2 factors), or high (all 3 factors). Weighted logistic regression models estimated the adjusted odds ratio (OR) of CVD for medium- and high-risk groups versus the low-risk group.

Results

CVD prevalence in medium-risk participants was not significantly different from low-risk participants (OR = 1.21, 95% CI: [0.86–1.70]). After adjusting for age, 

Contact: Thomas Karadimas <tkaradim@umich.edu>
Helen C.S. Meier

https://doi.org/10.3998/umurj.5512

159
high-risk participants had significantly higher odds of CVD prevalence compared to participants with 0 or 1 of the risk factors (OR = 1.86, 95% CI: [1.26–2.74]). This association was robust to additional demographic adjustment for sex, race/ethnicity, obesity status, smoking status, and diabetes mellitus or hyperglycemia.

**Conclusion**

Co-occurrence of hypertension, dyslipidemia, and elevated CRP was associated with CVD prevalence in a representative sample of older US adults. Our findings emphasize the importance of multifactor screening for CVD risk in clinical settings.

**Introduction**

Despite being one of the most researched and preventable chronic health conditions, cardiovascular disease (CVD) is the leading cause of death in the United States (US) and globally. CVD describes a range of conditions impacting the heart and vascular system, with the most common presentations being angina, myocardial infarction, and heart failure (CDC, 2018; Lopez et al., 2006; Flora & Nayak, 2019). The number of CVD cases has nearly doubled from 271 million in 1990 to 523 million in 2019, and the number of CVD-related deaths has also increased from 12.1 million in 1990 to 18.6 million in 2019 (Roth et al., 2020). In addition to contributing to morbidity and mortality globally, CVD has placed an extreme burden on healthcare systems and societies. In 2010, the global cost of CVD was an estimated $865 billion and is estimated to rise to $1.05 trillion by 2030. Around half of the monetary loss is due to direct healthcare costs, and the other half is productivity loss from not working, disability, or premature death (World Economic Forum, 2011). Therefore, the development of diagnostic tools for early identification of individuals at risk of developing CVD is crucial for timely diagnosis and intervention.

The cardiovascular system is made of the heart and blood vessels. CVD refers to the wide range of states of disease within this system (Farley et al., 2012). Although CVD can originate from a variety of elements, atherosclerosis, the obstructive deposition of fatty substances inside the arteries, is a primary cause of CVD (Libby et al., 2011).

The majority of CVD cases can be grouped into 4 categories:

1. **Coronary artery disease (CAD):** Plaque buildup in the arteries that supply blood to the heart, which leads to narrowing or blockage in these vessels. Results in angina, MI, and/or heart failure.
2. **Cerebrovascular disease (CVD):** Plaque buildup in the arteries that supply blood to the brain, which leads to narrowing or blockage in these vessels. Includes stroke and transient ischemic attack (TIA).
3. Peripheral artery disease (PAD): Plaque buildup in the peripheral arteries (located in the legs and arms), which leads to narrowing or blockage in these vessels.

4. Aortic atherosclerosis: Plaque buildup in the aorta, which leads to narrowing or blockage in this vessel, resulting in diminished blood flow and weakening of the aorta. Includes thoracic and abdominal aneurysms (Benjamin et al., 2018).

Multiple microscopic biological phenomena, including dyslipidemia, inflammation, and endothelial dysfunction, lead to the formation of fatty streaks (Davies et al., 1988), characteristic of atherosclerosis. Well-known risk factors contributing to these biomarkers of CVD are modifiable and include hypertension (HTN), dyslipidemia, diabetes mellitus (DM), smoking and secondhand smoke exposure, obesity, unhealthy diet, and sedentary lifestyle (CDC, 2022; Carey et al., 2021; Wilson et al., 1998; Alloubani, 2021; Jung et al., 2022). HTN is defined as a systolic blood pressure (SBP) ≥130mmHg and/or a diastolic blood pressure (DBP) >80mmHg and is the leading risk factor for CVD morbidity and mortality (Carey et al., 2021; Iqbal & Jamal, 2022). Dyslipidemia is defined as abnormal levels of lipids and/or lipoproteins in the blood, including elevated total cholesterol (TC ≥200 mg/dL), low-density lipoprotein cholesterol (LDL-C >100 mg/dL), triglyceride (TG >150 mg/dL), and low high-density lipoprotein cholesterol (HDL-C <40 mg/dL) (Alloubani, 2021, Jung et al., 2022). In the US, the number of adults with HTN is about 116 million and globally is about 4 billion (Roth et al., 2020). In the US, around 38% of adults have total cholesterol ≥200 mg/dL, 28% of adults have LDL ≥130 mg/dL, 21% of adults have TG ≥150 mg/dL, and 17% of adults have HDL <40 mg/dL (Tsao et al., 2022). Both HTN and dyslipidemia have been widely researched and have much data pointing toward their responsibility in the development of CVD, but studies have also demonstrated their insufficiency in independently producing all cases of CVD (Ridker et al., 2004; Libby et al., 2002; Khot, 2003; Johnson et al., 2018). Inflammation has emerged as another factor in the development of CVD (Sorriento & Iaccarino, 2019; Chen et al., 2022). Inflammation is the immune system’s reaction to dangerous substances, such as pathogens, damaged cells, toxins, or radiation. This process is crucial for maintaining health as it removes damaging entities and promotes healing. During acute inflammation, the immune system works to mitigate injury or infection. However, if inflammation becomes chronic, it can wreak havoc on the body and is known to contribute to a variety of debilitating health conditions, such as rheumatoid arthritis, lupus, periodontitis, and atopic dermatitis (Chen et al., 2017; Sorriento & Iaccarino, 2019). Past research has also shown a strong correlation between chronic inflammatory states and risk of developing atherosclerotic plaque and CVD (Koosha et al., 2020; Ferrucci &
In blood vessels, inflammation occurs in response to vessel injury, oxidation of serum lipids, and infection. HTN damages vessel endothelium, prompting an inflammatory response. In response to endothelial damage, leukocytes bind monocytes to the affected site. Dyslipidemia increases the risk of serum lipids getting stuck behind the endothelium and becoming oxidized. This triggers a stronger inflammatory response because monocytes that interact with these oxidized particles are more likely to stay bound to the endothelium. Furthermore, the monocytes can transform into macrophages and foam cells, which are precursors in the development of plaque. Other risk factors, including DM, smoking and secondhand smoke exposure, obesity, unhealthy diet, and sedentary lifestyle, amplify the harmful inflammatory effects of HTN and elevated lipids (Willerson & Ridker, 2004; Martinez-Quinones et al., 2018; Steinberg, 2009; Rhoads & Major, 2018; Tabas et al., 2007). Systemic inflammation can be measured by a variety of blood-based biomarkers, with the most prominent being C-reactive protein (CRP), which is strongly associated with CVD (Koosha et al., 2020; Ferrucci & Fabbri, 2018; Castro et al., 2018).

Although there is prior research examining the link between inflammation and CVD risk, there is limited research on the co-occurrence of inflammation, HTN, and dyslipidemia and resultant CVD risk. There remains an important question: Does an individual need all the 3 factors to develop CVD? Present research also includes samples that are often small and not directed toward geriatric populations, which is important considering older individuals are the primary group affected by CVD. In this present study, we will determine if having all 3 risk factors—high inflammation, hypertension, and dyslipidemia—correlates with a greater risk of CVD, as opposed to having 0–1 or 2 of these risk factors, using data from the Health and Retirement Study (HRS), a nationally representative, longitudinal study of adults over 50 years. We will describe the demographics (age, sex, race) and risk factors (obesity, smoking status, DM, HTN, dyslipidemia, inflammation) of HRS participants with CVD status. In addition, we will describe the demographics (age, sex, race) and risk factors (obesity, smoking status, DM, HTN, dyslipidemia, inflammation) of HRS participants with risk factor status. Finally, we will examine the association between risk factor status and CVD status, controlling for age, sex, race, obesity, smoking status, and DM.

Methods

Study Design and Data Sources

Data for this study come from HRS, an ongoing nationally representative panel study that surveys around 20,000 people older than age 50 in the US every
2 years. This study is supported by the National Institute on Aging (NIA) and Social Security Administration (SSA) and is coordinated by the University of Michigan Institute for Social Research (ISR). The design and methodology of HRS have been thoroughly outlined elsewhere (Sonnega et al., 2014; Karp, 2007; Juster & Suzman, 1993). In this study, we carried out a cross-sectional analysis examining the association between CVD prevalence and clinical risk factors for CVD using data from 2016 (Wave 13). Of the 9,850 participants with biomarker data available at this wave, we excluded individuals missing data on BP, lipids, and CRP, as well as covariates including age, sex, race/ethnicity, DM or hyperglycemia, smoking, and obesity to obtain a final sample of 1,527 participants.

**Exposure**

CVD risk score was generated from 3 clinical conditions known to contribute to CVD: HTN, dyslipidemia, and elevated CRP. HTN status was obtained from HRS self-reported health data on the basis of a healthcare provider’s diagnosis. LDL-C was calculated in serum specimens having a triglyceride value <400 mg/dL using the formula of Friedewald LDL-C = TC – HDL-C – TG/5.0. HDL-C was measured directly in serum using the Roche HDL-C 3rd-generation direct method (Roche Diagnostics, Indianapolis, IN) on a Roche Cobas 6000 Chemistry Analyzer (Roche Diagnostics Corporation). TG was measured in serum using an enzymatic TG Reagent (Roche Diagnostics, Indianapolis, IN) on a Roche Cobas 6000 Chemistry Analyzer (Roche Diagnostics Corporation). Dyslipidemia was categorized as LDL-C >130 mg/dL, TG >150 mg/dL, or HDL-C <50 mg/dL for women and <40 mg/dL for men, based on widely recognized clinical reference ranges (Martin & Cardoso, 2021). CRP was measured in serum using a latex-particle enhanced immunoturbidimetric assay kit (Roche Diagnostics, Indianapolis, IN 46250) and read on the Roche COBAS 6000 Chemistry analyzer (Roche Diagnostics). High CRP was categorized as >3 mg/L, based on the widely recognized clinical reference range (HRS, 2016; Adukauskienė et al., 2016). Values of continuous CRP were log transformed for descriptive statistics.

Low risk was defined as having 0–1 diagnoses of HTN, high CRP, or dyslipidemia; medium risk was defined as having 2 of these factors; and high risk was defined as having all 3 factors.

**Outcome**

Self-reported health data of a healthcare provider’s diagnosis of heart attack, coronary heart disease, angina, congestive heart failure or other heart problems,
and stroke or transient ischemic attack (TIA) were used to assess prevalence of CVD. If a participant reported either a heart condition or a stroke or TIA, they were classified as having CVD.

**Covariates**

Covariates included age (years), sex assigned at birth (female/male), race/ethnicity (non-Hispanic white, non-Hispanic black, or Hispanic), obesity (yes/no), current smoking status (yes/no), and DM or hyperglycemia (yes/no). Age, sex, and race/ethnicity were obtained from HRS self-reported demographic data. The presence of DM or hyperglycemia was obtained from HRS self-reported health data on the basis of a healthcare provider’s diagnosis. Smoking status was obtained from HRS self-reported health data. A participant was classified as obese if their calculated body mass index was ≥30 kg/m², based on the widely recognized clinical classification of obesity (Jensen et al., 2014).

**Statistical Analysis**

Continuous variables were reported as weighted mean ± standard error, and categorical variables were reported as frequency (weighted percentage). Distributions of covariates by CVD diagnosis and risk score were reported using F-test for continuous variables and Rao-Scott Chi-Square Test for categorical variables. Weighted logistic regression models were used to calculate adjusted odds ratio (OR) of CVD in participants with low-, medium-, and high-risk groups. Model 1 was adjusted for age. Model 2 was adjusted for age, sex, race, obesity, smoking status, and DM.

All analyses were completed using SAS version 9.4 (SAS Institute, Inc., Cary, NC). P values at alpha < 0.05 were considered statistically significant.

**Results**

Descriptive statistics for those with and without CVD are provided in Table 1. The mean age of participants with CVD was 70.98 ± 0.56 years and 66.65 ± 0.33 years for participants without CVD (P < 0.05). Of the participants with CVD, 47.79% were female, and of the participants without CVD, 59.93% were female (P < 0.05). Distributions of race/ethnicity, obesity status, and smoking status were not significantly different between the two groups. There was a greater prevalence of DM or hyperglycemia in the CVD group. Mean LDL-C and HDL-C levels were
significantly greater in the no CVD group, while there were no significant differences in mean TG or log CRP between the CVD and no CVD groups. HTN and high CRP were more prevalent in the CVD group, while dyslipidemia was more prevalent in the no CVD group.

Distributions of covariates for low-, medium-, and high-risk factor scores based on presence of 0–1, 2, or 3 of HTN, dyslipidemia, and high CRP, respectively, are described in Table 2. The low-risk group mean age was 68.18 ± 0.49 years, the medium-risk group mean age was 67.70 ± 0.43 years, and the high-risk group mean age was 67.62 ± 0.60 years \((P > 0.05)\). About 53.66%, 59.59%, and 56.17% of participants in the low-, medium-, and high-risk groups were female.

<table>
<thead>
<tr>
<th>Variables</th>
<th>CVD ((n = 470))</th>
<th>No CVD ((n = 1057))</th>
<th>(P) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>70.98 ± 0.56</td>
<td>66.65 ± 0.33</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>246 (47.79)</td>
<td>653 (59.93)</td>
<td>0.0006</td>
</tr>
<tr>
<td>Race/Ethnicity, n (%)</td>
<td>Non-Hispanic white 336 (82.30)</td>
<td>691 (81.19)</td>
<td>0.8809</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic black 75 (8.82)</td>
<td>188 (9.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hispanic         59 (8.88)</td>
<td>178 (9.78)</td>
<td></td>
</tr>
<tr>
<td>Obesity, n (%)</td>
<td>223 (47.30)</td>
<td>461 (41.78)</td>
<td>0.1211</td>
</tr>
<tr>
<td>Current Smoker, n (%)</td>
<td>50 (8.82)</td>
<td>110 (11.11)</td>
<td>0.2768</td>
</tr>
<tr>
<td>DM or Hyperglycemia, n (%)</td>
<td>182 (36.07)</td>
<td>251 (20.25)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LDL-C (mg/dL)</td>
<td>94.26 ± 2.09</td>
<td>108.76 ± 1.38</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HDL-C (mg/dL)</td>
<td>51.97 ± 0.95</td>
<td>60.42 ± 0.81</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TG (mg/dL)</td>
<td>139.44 ± 3.98</td>
<td>137.27 ± 2.68</td>
<td>0.6503</td>
</tr>
<tr>
<td>Log CRP</td>
<td>0.75 ± 0.07</td>
<td>0.59 ± 0.05</td>
<td>0.0698</td>
</tr>
<tr>
<td>HTN, n (%)</td>
<td>375 (76.05)</td>
<td>622 (54.42)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Dyslipidemia, n (%)</td>
<td>301 (65.19)</td>
<td>795 (76.04)</td>
<td>0.0006</td>
</tr>
<tr>
<td>High CRP, n (%)</td>
<td>211 (42.27)</td>
<td>405 (35.41)</td>
<td>0.0470</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of HRS Participants with and without CVD \((n = 1527)\)

\(n\): number of participants, LDL-C: low-density lipoprotein cholesterol, mg/dL: milligrams per deciliter, HDL-C: high-density lipoprotein cholesterol, TG: triglyceride, CRP: C-reactive protein, HTN: hypertension.

Values are mean ± standard error and frequency (percentage); **Bold** = statistically significant at alpha < 0.05.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Low-Risk ($n = 582$)</th>
<th>Medium-Risk ($n = 601$)</th>
<th>High-Risk ($n = 344$)</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>68.18 ± 0.49</td>
<td>67.70 ± 0.43</td>
<td>67.62 ± 0.60</td>
<td>0.7004</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>322 (53.66)</td>
<td>364 (59.59)</td>
<td>213 (56.17)</td>
<td>0.2747</td>
</tr>
<tr>
<td>Race/Ethnicity, n (%)</td>
<td>Non-Hispanic white</td>
<td>436 (86.18)</td>
<td>394 (80.09)</td>
<td>197 (74.87)</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic black</td>
<td>75 (6.73)</td>
<td>101 (8.39)</td>
<td>87 (14.53)</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>71 (7.09)</td>
<td>106 (11.53)</td>
<td>60 (10.60)</td>
</tr>
<tr>
<td>Obesity, n (%)</td>
<td>164 (25.67)</td>
<td>296 (48.42)</td>
<td>224 (68.95)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Current Smoker, n (%)</td>
<td>57 (8.64)</td>
<td>56 (10.08)</td>
<td>47 (14.79)</td>
<td>0.0910</td>
</tr>
<tr>
<td>DM or Hyperglycemia, n (%)</td>
<td>132 (19.32)</td>
<td>171 (25.43)</td>
<td>130 (34.32)</td>
<td>0.0003</td>
</tr>
<tr>
<td>LDL-C (mg/dL)</td>
<td>96.31 ± 1.70</td>
<td>109.38 ± 1.83</td>
<td>112.21 ± 2.78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HDL-C (mg/dL)</td>
<td>64.61 ± 1.08</td>
<td>56.01 ± 0.97</td>
<td>48.66 ± 1.04</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TG (mg/dL)</td>
<td>114.88 ± 3.02</td>
<td>146.62 ± 3.33</td>
<td>167.17 ± 5.46</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Log CRP</td>
<td>0.09 ± 0.05</td>
<td>0.68 ± 0.06</td>
<td>1.66 ± 0.07</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HTN, n (%)</td>
<td>193 (28.11)</td>
<td>460 (73.91)</td>
<td>344 (100.00)</td>
<td></td>
</tr>
<tr>
<td>Dyslipidemia, n (%)</td>
<td>249 (46.08)</td>
<td>503 (86.92)</td>
<td>344 (100.00)</td>
<td></td>
</tr>
<tr>
<td>High CRP, n (%)</td>
<td>33 (4.17)</td>
<td>239 (39.17)</td>
<td>344 (100.00)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Characteristics of HRS Participants by Risk Category ($n = 1527$)

$n$: number of participants, LDL-C: low-density lipoprotein cholesterol, mg/dL: milligrams per deciliter, HDL-C: high-density lipoprotein cholesterol, TG: triglyceride, CRP: C-reactive protein, HTN: hypertension.

Values are mean ± standard error and frequency (percentage); **Bold** = statistically significant at $alpha < 0.05$. 

UMURJ • vol. 17, no. 1 • 2024
respectively ($P > 0.05$). The proportion of non-Hispanic whites was highest in the low-risk group, while the medium and high-risk groups contained a greater proportion of non-Hispanic black participants and Hispanic participants compared to the low-risk group. The prevalence of obesity was patterned according to the risk group, with the highest average BMI in the high-risk group and lowest in the low-risk group. Smoking status was not significantly different between the three groups. There was an increasing prevalence of DM or hyperglycemia as the number of risk factors increased. Mean LDL-C, TG, and log CRP also increased with the increasing risk factor score, while mean HDL-C decreased.

Associations between risk factor score and CVD prevalence are shown in Table 3 and Figure 1. The prevalence of CVD in medium-risk participants was

<table>
<thead>
<tr>
<th></th>
<th>Low-Risk</th>
<th>Medium-Risk</th>
<th>High-Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Ref</td>
<td>1.21 (0.86–1.70)</td>
<td>1.86 (1.26–2.74)</td>
</tr>
<tr>
<td>Model 2</td>
<td>Ref</td>
<td>1.19 (0.84–1.68)</td>
<td>1.68 (1.12–2.53)</td>
</tr>
</tbody>
</table>

**Table 3:** Adjusted Odds of CVD Prevalence for Risk Category in HRS Participants ($n = 1527$)

OR: odds ratio, CI: confidence interval.
Model 1: Adjusted for age.
Model 2: Adjusted for age, sex, race/ethnicity, obesity, current smoker, DM, or hyperglycemia.

**Figure 1:** Association (Odds Ratios and 95% Confidence Intervals) between Risk Category and CVD Prevalence in HRS Participants ($n = 1527$)
not significantly different compared to low-risk participants. After adjusting for age, the odds of CVD prevalence in high-risk participants were 1.86-fold higher (95% CI: [1.26–2.74]) compared to low-risk participants. This association was robust to additional adjustment for sex, race/ethnicity, obesity status, smoking status, and DM or hyperglycemia status (OR = 1.68, 95% CI: [1.12–2.53]).

Discussion

HTN and dyslipidemia are well established as contributing factors in the development of CVD. More recently, elevated inflammatory status, measured through elevated CRP, has been explored as an important element in determining the risk of CVD. This study investigated the prevalence of CVD between groups with 0–1, 2, or 3 of these conditions in a nationally representative sample of US adults over the age of 50.

Our findings revealed that those with 0–1 or 2 risk factors all had similar prevalence of CVD, but those with all 3 CVD risk factors had a significantly higher prevalence of CVD, even after adjusting for age, sex, race/ethnicity, obesity status, smoking status, and presence of DM or hyperglycemia. This supports the hypothesis that CVD is multifactorial, and the combination of HTN, dyslipidemia, and inflammation is characteristic of those with healthcare provider–diagnosed CVD. As a result, screening for all 3 factors may be clinically important for identifying those at the risk of developing CVD.

These results are consistent with previous research, including multiple observational and experimental studies. Previously, based on the traditional cholesterol hypothesis of CVD, which places the blame of atherosclerosis primarily on cholesterol, many medications were produced with the goal of reducing cholesterol. One promising medication, niacin, resulted in significant improvements in participant lipid profiles, including lowered TG and LDL-C and greater HDL-C. Despite these improvements in lipids, participants demonstrated zero clinical benefit, and some even had worse CVD outcomes (AIM-HIGH Investigators, 2011). In addition, even on multidrug lipid-lowering therapy, 1 in 5 participants experienced residual CVD risk (Alfaddagh et al., 2020; Sampson et al., 2011). In multiple CVD drug trials, including PROVE-IT and IMPROVE-IT, residual CVD risk was associated with CRP elevation despite achieving control of cholesterol (Ridker et al., 2005; Bohula et al., 2015). Other CVD drug trials, such as JUPITER and CANTOS, have found that the anti-inflammatory and CRP-lowering effects of drugs, including statins, produce a protective effect against CVD independent of cholesterol-lowering effects.
Association between Coexisting Hypertension, Dyslipidemia, and Elevated

(Willerson & Ridker, 2004; Ridker et al., 2009a; Ridker et al., 2009b; Ridker et al., 2017; Ridker et al., 2023). Our current knowledge has moved from believing that CVD stemmed from cholesterol buildup alone to a multifaceted disease, cholesterol being just one piece of the ever-changing puzzle that is the human body. HTN and dyslipidemia are the most significant players in the development of CVD, but studies have started to point toward inflammation as a component in the etiology of CVD as well. Inflammation is likely to be involved in atherosclerosis, and determining CVD risk based on a plethora of factors, not just BP and cholesterol, is prudent in the mitigation of CVD-related deaths. In practice, proper and comprehensive assessment of BP, lipids, and CRP is rare, which is especially problematic in CVD, a multifactorial condition. In this context, it is important to ensure implementation of current diagnostic guidelines, which advocate for concurrent and routine BP and lipid measurements in all adults, as well as CRP measurements in individuals at the risk of CVD, because just knowing one or two of these health parameters is not enough to accurately describe CVD risk (Pearson et al., 2003).

Overall, participants with CVD tended to be older and were more likely to be male and to have DM or hyperglycemia, as expected. In addition, participants with CVD had lower LDL-C and HDL-C levels, were more likely to have HTN and elevated CRP, and were less likely to be dyslipidemic. The lower prevalence of dyslipidemia in those with CVD could be due to lipid-lowering medication use, which is a first-line agent in the prevention of future CVD events in those previously diagnosed with CVD. Thorough analysis of medication usage is beyond the scope of this paper but is an area for future research. We found that non-Hispanic black and Hispanic participants made up a greater proportion of the high-risk category than the low-risk category, and vice versa, for non-Hispanic white participants. In addition, high-risk participants were more likely to be obese and to have DM or hyperglycemia.

This study has many strengths. Data come from HRS, a large and representative sample of older US adults. Measures of key laboratory variables, including CRP and lipids, were obtained using a standard protocol. HRS contains a wealth of information on covariates, allowing the analyses to be controlled for potential confounders of the association. Limitations of this study include the absence of longitudinal hypertensive, lipid, inflammatory, and cardiovascular data. An analysis of this longitudinal data would provide insight into the temporal order of CVD risk factors and CVD and thus can support causal relationships between these variables. In addition, CVD, obesity, smoking status, DM or hyperglycemia, and HTN are all self-reported from participants, and these data can be influenced by memory, social desirability bias, and absence of diagnosis. Smoking status and obesity are highly variable across individuals, and with a yes/no
metric, it is difficult to ascertain the severity of these variables (i.e., how many pack-years and overall BMI). Other lifestyle information, such as physical activity, diet, and stress, which are all independent factors in development of CVD (Osborne et al., 2020; Li & Siegrist, 2012; Casas et al., 2018), was not included as covariates. Furthermore, pharmacological interventions are not analyzed, as it is outside the scope of this study. Finally, biomarkers that have been shown to more accurately predict CVD risk, such as LDL particle size, Lipoprotein(a), and Apolipoprotein B, were not available for analysis in HRS (Lamarche et al., 1997; Behbodikhah et al., 2021; Duarte Lau & Giugliano, 2022). Future studies should focus on implementing a variety of lifestyle factors and above-mentioned biomarkers on top of the CVD risk factors examined in this study to develop a comprehensive score of CVD risk.

Conclusion

The co-occurrence of HTN, dyslipidemia, and elevated CRP was associated with CVD prevalence in a representative sample of older US adults. Our findings emphasize the importance of multifactor screening for assessing CVD risk in the clinical setting.

Abbreviations

CVD: cardiovascular disease
US: United States
CAD: coronary artery disease
CHD: coronary heart disease
PAD: peripheral artery disease
HTN: hypertension
DM: diabetes mellitus
SBP: systolic blood pressure
DBP: diastolic blood pressure
mmHg: millimeters of mercury
TC: total cholesterol
LDL-C: low-density lipoprotein cholesterol
HDL-C: high-density lipoprotein cholesterol
TG: triglyceride
CRP: C-reactive protein
Association between Coexisting Hypertension, Dyslipidemia, and Elevated

\[ n: \] number of participants

\[ \text{mg/dL}: \] milligrams per deciliter

HRS: Health and Retirement Study

NIA: National Institute on Aging

ISR: Institute for Social Research

SSA: Social Security Administration

TC: total cholesterol

OR: odds ratio

CI: confidence interval

References


Association between Coexisting Hypertension, Dyslipidemia, and Elevated


Fitness Costs of Early Life Adversity in Non-Human Primates

IRENE WANG

As arguably the most critical period of mammalian development, early life can be defined as the period from conception to reproductive maturity (Tung et al., 2016). The importance of adequate nutrition and proper maternal care for maximizing the survival and development of offspring has been demonstrated through research findings in various species (Lu et al., 2018). However, these ideal growth conditions are difficult to achieve for wild offspring, who often have to overcome various energetically or socially challenging conditions that limit their development in early life. This paper will focus on the effect that environmental and social adversities have on the fitness of wild and captive non-human primates as measured by lifespan and reproductive success. Current research suggests that early life adversity in non-human primates reduces individual fitness by decreasing lifespan, fertility, and offspring survival. This paper explores and discusses the various factors that contribute to this fitness reduction. These factors include but are not limited to increased risk of physical illness and behavioral disorders, social isolation, maternal death, intergenerational effects, reduced female fertility, maternal stress, delayed sexual maturation, and lower adult body size.

Introduction

Early life adversity can affect important developmental outcomes, ranging from social behavior, gene regulation and expression, and cognition and neuroanatomically functions (French & Carp, 2016). Causes of early life adversity can be broadly classified into two categories: environmental (e.g., resource limitation, predation, disease, and natural disasters) and social (e.g., dominance rank, maternal affiliative social connectedness, maternal death, and sibling competition). The main mechanism through which early life adversity modifies an individual’s development is through adaptive plasticity, defined as the ability of

Contact: Irene Wang <irenew@umich.edu>
genetically similar individuals to develop potentially adaptive phenotypic differences in response to different early life experiences (Lu et al., 2018). Given the profound impact early life adversity has on offspring survival and development, it is important to understand the underlying developmental and evolutionary mechanisms behind adaptive plasticity.

There are two major competing views regarding the fitness cost-benefit analysis of early life adversity and adaptive plasticity: the developmental constraints hypothesis and the predictive adaptive response (PAR) hypothesis. The developmental constraints hypothesis states that offspring born in adverse early environments are at a fitness disadvantage relative to their conspecifics, regardless of adult environmental conditions (Lea et al., 2015). Under this hypothesis, short-term adaptive changes to adverse early life conditions come with long-term survival and reproductive costs. The PAR hypothesis, on the contrary, states that adverse early environments prompt adaptive phenotypic adjustments that prepare animals for similar challenges in adulthood (Lea et al. 2015). This suggests that early life adversity may bring fitness advantages to individuals by preparing them for an equally adverse adult environment. Current research in long-lived species, such as primates, appears to support the developmental constraint model where individuals raised in adverse conditions suffer lifelong fitness disadvantages compared to their conspecifics regardless of their adult environment. Offspring born and raised in adverse environments experience higher mortality rates and reduced ability to care for their own offspring (Zipple et al., 2019). The research data and results that support the developmental constraints hypothesis are discussed next.

**Reduced Lifespan**

An individual’s lifespan can serve as a reliable indicator of their fitness. This is especially true in most primate species where there is little to no post-reproductive period. Tung et al. (2016) found there are strong correlations between an individual’s age at death and total number of offspring who survived to at least age 1. Individuals with longer lifespans will have more opportunities to conceive and time to raise their offspring. In wild female baboons, those who experience three or more sources of early adversity die on average 10 years earlier than females who experience little to no adverse conditions (Tung et al., 2016). This is consistent regardless of the match between early life and adult environmental conditions. Their significantly shortened lifespan is likely caused by early life malnourishment and an accumulation of environmental aggravations that steadily weaken their physical conditions (Tung et al., 2016).

The detrimental effects accumulated through early life are seldom reversed in adulthood despite adequate resource availability (Lea et al., 2015). Offspring
raised in adverse conditions often have low-ranking mothers whose own physical condition is relatively poor (Zipple et al., 2020). There is little chance for the offspring to move up the social hierarchy, as their weakened condition significantly reduces their resource competition ability. In wild baboons, females who experience the most adversity are also more socially isolated as adults (Tung et al., 2016). Strong social bonds are a known factor associated with increased reproductive success and longevity in baboons (Silk et al., 2010). This suggests that reduced social bond strength experienced by high adversity females can have a negative impact on their physical health and their lifespan.

Higher Prevalence of Physical Illness and Behavioral Disorders

One of the more obvious ways in which early life adversity can reduce fitness is by increasing the offspring’s susceptibility to both physical illnesses and behavioral disorders.

In experimental studies conducted by Conti et al. (2012), captive rhesus macaques raised by either humans or surrogate mothers (warm water bottles) have higher rates of developing physical illnesses than conspecifics raised by their biological mothers. Behavioral stereotypies, such as inappropriate vocalizations, are also more prevalent in these individuals (Conti et al., 2012). Nursery-reared macaques also demonstrate greater behavioral inhibition and abnormal social affective behavior compared to their mother-reared counterparts in adulthood (Corcoran et al., 2011). These behavioral deviations can be detrimental to their social relationships with other macaques as adults and lead to isolation. Since macaques are a very social species, this social isolation can lead to poor physical health, much like the effects seen in baboons (Corcoran et al., 2011). Social isolation can also heighten individual susceptibility to predation and reduce ability to obtain resources, both of which can significantly decrease the individual’s lifespan and reproductive success (Tung et al., 2016).

Maternal Death

Most mammalian species rely on their mother for food, safety, and learning important survival skills during early life. It is predicted, then, that maternal death during this period would have a significant impact on offspring survival. Indeed, adult female baboons who lost their mothers prior to maturity were three times more likely to die than their conspecifics (Tung et al., 2016). Similar results have been found in multiple wild Old World and New World primate species (Zipple et al., 2020). However, offspring survival is reduced well
before the mother faces imminent death. This can be explained by the maternal
conditions hypothesis, which states that mothers who are in poor physical con-
ditions are more likely to die. Thus, imminent maternal death can serve as a
proxy for poor maternal condition, which can then result in decreased ability to
provide and care for offspring (Zipple et al., 2020). In addition, female offspring
who experience early maternal death have reduced ability to produce and suc-
cessfully rear their own offspring to maturity in adulthood (Zipple et al., 2020).
These females likely have weakened physical conditions that make gestation
and lactation highly taxing on their physical health as it severely interferes with
their resource-gathering abilities. This suggests that early adversity experienced
in the form of maternal death may have intergenerational effects.

Intergenerational Effects of Early Life Adversity

The intergenerational and long-term effects of early life adversity have been over-
looked by previous literature for some time due to a lack of long-term data. As
cumulative data becomes available in recent years, studies have revealed some
striking finds regarding the intergenerational effects of early life adversity in
wild baboons. Offspring who did not experience early adversity of their own but
whose mother experienced early adversity exhibit a 48% higher mortality rate in
the first four years of life than their unaffected conspecifics (Zipple et al., 2019).
Females who experienced early adversity also demonstrated reduced ability to
care for their own offspring near the end of life compared to their conspecifics.
The reduction in offspring rearing ability could be the result of these females’
own poor physical condition as a consequence of early adversity. It could also
be due to a lack of offspring rearing skills, as they may have little opportunity to
observe and practice these skills if their mothers died early. This is a classic exam-
ple of the parental effect, where the environment experienced by an individual
can affect their own phenotype and thus their offspring’s phenotype (Zipple et al.,
2019). This then forms a vicious cycle where one’s present predicament as a result
of early adversity causes one’s offspring to suffer through the same. The exact
mechanism for this intergenerational effect is still under investigation. Based on
these findings, it is clear that early life adversity can bring long-term reductions
on both the direct and the indirect fitness of offspring across generations.

Reduced Fertility in Wild Female Amboseli Baboons

Early life adversity can also have a direct impact on reproductive success by
reducing fertility during extreme adverse conditions in adulthood. Long-term
field data from the Amboseli Baboon Research Project in Kenya indicates that females born in low-quality environments showed greater decrease in fertility during drought years than females born in high-quality environments. This reduction in fertility is associated with reduction in these females’ fitness since they produce less offspring than conspecifics. This directly opposes the PAR hypothesis, since it is not the case that females whose early life conditions matched with adulthood conditions gain a fitness advantage. Instead, they are at a disadvantage during adverse conditions compared to conspecifics. Interestingly, females who are reared by high-ranking mothers during adverse environmental conditions did not experience this reduction in fertility. In fact, each improvement of one maternal rank position would increase resumption of cycling probability and conception probability during severe drought years by 1.8% and 1.3%, respectively (Lea et al., 2015). This suggests that adequate resource provisioning acquired through high maternal social status and support can buffer offspring against adverse natural conditions.

Reduced fertility may be an adaptive reproductive trade-off where females suspend reproduction in order to invest more attention and resources to their current offspring during adverse conditions. Yet, research data in savannah baboons suggests that this increase in maternal investment by low-ranking females did not increase offspring size or survival (Altmann & Alberts, 2005). Mothers who experienced early life adversity often have very limited resource access and poor physical conditions (Altmann & Alberts, 2005). These significant limitations to their ability for offspring care simply cannot be compensated through effort alone.

Maternal Stress

High-ranking female baboons and those with high affiliative social connectedness also experience lower maternal stress (Lea et al., 2015). This is likely due to the affiliative gestures such as reciprocal grooming and possible protection from predators and infanticide that one’s social partners may provide. Glucocorticoids are a class of steroid hormones linked with various inflammatory and immune responses. The reduction in stress and relatively stable glucocorticoid levels has been linked with stronger immune defense and lower inflammatory response in baboons (Campos et al., 2021). Chronic stress and high glucocorticoid levels experienced by early adversity offspring are linked with significant reductions in lifespan by as much as 5.4 years (Campos et al., 2021). Offspring raised by mothers who are under chronic stress may be neglected and experience elevated stress levels as well, further perpetuating the cycle of affective instability. Studies that focus on neuroendocrinology involving glucocorticoids and gut
microbiome are underway to explore the epigenetic and other biological changes that can affect an individual’s long-term health outcomes (Lu et al., 2018). They serve as practical measures of physical attributes that can give important insight into an individual’s physical condition and well-being. Understanding these physical effects of early life adversity can provide clues to the biological mechanisms by which early life adversity predicts adult fitness.

**Delayed Sexual Maturation and Reduced Adult Body Size**

Delayed sexual maturation and smaller adult body size observed in savannah baboons raised in resource-limited environments counter the PAR hypothesis. Altmann and Alberts (2004) found that offspring raised in resource-rich environments attained accelerated growth rate and earlier sexual maturation than their resource-limited conspecifics. They also achieved a larger relative and absolute body size than their counterparts (Altmann & Alberts, 2004). According to the PAR hypothesis, the opposite pattern should be observed where, in order to compensate for reduced expected lifespan suggested by environmental cues in early life, offspring that experience early life adversity should have a faster sexual maturation rate so that they can produce offspring earlier than their conspecifics. Yet, the research data directly counters this prediction. Once again, offspring who experience early life adversity are at a fitness disadvantage in terms of growth and reproductive success.

**Other Considerations**

**Maternal Capital Model**

Both the developmental constraints and PAR models place their emphasis on offspring survival and fitness. The maternal capital model, however, focuses on the survival of the mother and the decisions and strategies employed to maximize her own fitness (Lu et al., 2018). This model suggests that developmental processes are designed to increase maternal rather than offspring fitness. Offspring receive environmental information indirectly through their mother, who may filter this information to suit her own fitness needs (Lu et al., 2018). Major developmental decisions, such as weaning age and resource allocation, are based on the mother’s needs rather than the offspring’s. This suggests that offspring development is mainly dependent on their mother’s fitness cost-benefit analysis. More attention should be given to test and investigate the extent of maternal control on offspring development trajectories.
Sibling Competition

Sibling competition is a social factor that could be considered as an adversity since it increases resource and maternal capital competition of the offspring. Research found that female baboons with close-in-age younger siblings were twice more likely to die at every age than females without siblings (Tung, 2016). Another study in wild baboons by Zipple et al. mirrored this result in intergenerational competition where offspring whose mother has close-in-age siblings experience 39% higher mortality than their counterparts (Zipple et al., 2019). Sibling competition is well documented in birds but is currently lacking in primate studies. It can have a powerful effect on offspring development, especially considering the slow life histories of primates. Given that gestation and lactation are incredibly energetically costly processes for mothers, having multiple offspring in short birth intervals can severely exhaust the mother’s body and reduce her foraging ability. It can be very difficult to keep up with the demands of two young offspring at once, both resource and attention wise. This might lead to severe reduction in offspring fitness since adequate care for one comes at a cost for their siblings.

Sex Differences

It is noteworthy that many of the studies mentioned in this paper focused on only one sex, usually females. This paints an incomplete picture to the understanding of early life adversity’s impact on individual survival and fitness, as different sexes often operate under different social conditions and reproductive strategies. There is conflicting evidence as to the significance of differential effects of sex when it comes to early life adversity in primates (French & Carp, 2016). Future studies should consider both sexes when analyzing research data regarding early life adversity. The effects of such early life adversities can be far-reaching and negatively impact the individuals for the rest of their life.

Conclusion

Early life adversity can have profound impacts on the long-term health and reproductive success of the individual. The exact evolutionary and biological mechanisms by which early life adversity operates to manifest these impacts is currently not fully understood. The two leading theories of developmental constraints hypothesis and PAR hypothesis, plus newer hypotheses such as the
maternal capital model, provide guiding research frameworks that are testable both in the lab and out in the field.

Non-human primates are human’s closest living relatives. Given their biological relatedness and social similarities to humans, non-human primate models can provide insightful clues regarding human growth and development. Many of the adverse conditions in these studies cannot be replicated or experimented on humans as it would raise serious ethical concerns. Therefore, primate studies are the closest route through which human developmental theories can be tested. Experimental data gathered from these studies can be used to develop and improve existing models regarding not only human development but also evolutionary mechanisms of adaptive plasticity in other species as well.

There is still much left to be explored regarding the role of maternal care, environmental conditions, and social adversity in offspring’s physical and social development. Since research in this area relies heavily on long-term data, the exciting potential of advances in the field is just starting to be unveiled. Given that the effects of such early life adversities can be far-reaching and negatively impact the individual for the rest of their life, future research should remain conscious of the long-term implications that experiments may have for the test subjects and implement humane mitigations where possible.

References


