



Perceptions, Health-Seeking Behaviour and Traditional Treatment of Common Childhood Illnesses in Yemen: A Qualitative Study

Table of Contents

3 Editorial

Abdulrazak Abyad
Chief Editor

Original Contribution

5 Perceptions, Health-Seeking Behaviour and Traditional Treatment of Common Childhood Illnesses in Yemen: A Qualitative Study

Fauzia Faraj Bamatraf
DOI: 10.5742/MEWFM.2026.241792

Population and Community Studies

18 Health System Expansion Without Proportional Health Gains? An Open-Data Policy Analysis of Healthcare Utilization and Overuse in Türkiye

Hakan Yaman
DOI: 10.5742/MEWFM.2026.241793

Dementia Special Series

36 Cognitive Testing: Bedside Screening Tools and Formal Neuropsychological Assessment - A Comprehensive Academic Review

A Abyad
DOI: 10.5742/MEWFM.2026.241794

Review

55 Neutropenia in Infants: A Comprehensive Narrative Review

Khawla Faraq Alzaitoni
DOI: 10.5742/MEWFM.2026.241796

Education and Training

66 Invisible Curriculum, Unmeasured Outcomes: A Theory-Informed Analysis of Quality Assurance Gaps in Family Medicine Training

Hakan Yaman, Aylin Yaman
DOI: 10.5742/MEWFM.2026.241798

Letter to the Editor

76 Data rich, meaning poor

Hakan Yaman

Editorial

The first paper in this issue directs attention to decisions that often precede a clinical consultation: how families recognise illness, explain its causes, choose treatment, and decide when professional care is necessary. In *Perceptions, Health-Seeking Behavior and Traditional Treatment of Common Childhood Illnesses in Yemen: A Qualitative Study*, Fauzia Faraj Bamatraf examines these decisions among caregivers in Al-Mukalla. Her findings deserve particular attention because they identify opportunities for family physicians to address delays in care through better communication, safer home management, and closer engagement with the people who influence a child's treatment.

Conducted between January and May 2025, the study involved three focus groups with 21 caregivers of children under five, together with interviews with five traditional healers and observations of treatment practices. Participants were purposively selected, and caregivers had previous experience using traditional medicine or home remedies. This approach provides insight into how these participants understood childhood illness and acted on that understanding. It does not estimate the prevalence of traditional medicine use across Yemen or establish the effects of particular practices on mortality. Its contribution is to explain decisions that routine service statistics cannot adequately describe.

A central finding is that caregivers' explanations of illness shaped which treatments they considered appropriate. Participants sometimes understood diarrhoea, vomiting, or other symptoms as part of teething and normal development. Other explanations included accidental falls, the evil eye, or the perceived unsuitability of breast milk. Biomedical explanations, including poor hygiene and contaminated food, existed alongside these beliefs. Some caregivers regarded particular conditions as illnesses for which medical treatment would be ineffective and traditional treatment

necessary. Under these circumstances, knowing where a clinic is located does not necessarily lead a family to seek care there.

This distinction has practical importance. Health education that merely advises families to consult a doctor may leave the underlying reason for delay untouched. If a caregiver believes an illness cannot respond to medical treatment, repeating instructions without discussing that belief is unlikely to resolve the disagreement. Bamatraf's findings support a more attentive consultation in which clinicians ask what the family thinks has caused the illness, which treatments have already been tried, and what changes would prompt them to seek further help. Such questions make the family's reasoning available for discussion and allow advice to address the actual concern.

The study also shows that treatment decisions were shared within families and communities. Mothers consulted grandmothers, mothers-in-law, friends, and neighbours. Traditional healers offered familiar explanations and treatments that participants described as accessible and affordable. These accounts suggest that educational programmes should include influential family members and community figures. Advice given to one caregiver may have limited effect if other decision-makers continue to recommend a different course of action. Family medicine is well placed to recognise these relationships and incorporate them into counselling.

Care is needed, however, to avoid explaining delayed treatment solely through culture. Participants also described financial constraints and the convenience of home remedies. One account linked hospital attendance to the availability of money, with family-provided remedies used in the meantime. The appropriate response therefore extends beyond correcting misconceptions. Community education should be accompanied by practical attention to the cost of care, availability of treatment, and the ability to reach a functioning service. Families need recommendations they can follow.

Several reported practices warrant direct clinical attention. Caregivers described withholding breastfeeding during diarrhoeal illness, using leftover or shared medicines, and administering antibiotics without a clearly identified prescription. Only a few mentioned oral rehydration solution. These findings identify specific subjects for counselling: maintaining feeding, demonstrating the correct preparation and use of oral rehydration solution, and reviewing all medicines and remedies given to a child. WHO guidance supports oral rehydration and continued nutrient-rich feeding, including breast milk, during diarrhoeal illness. WHO, *Diarrhoeal disease*

The reported use of honey from birth also requires an explicit safety message. Honey should not be given to infants younger than 12 months because of the risk of infant botulism. Its traditional reputation or use in other clinical contexts does not establish its safety when fed to an infant. This is precisely the kind of specific, actionable information that should accompany respectful discussion of home remedies. CDC, *Foods and drinks to avoid or limit*.

The observations of cauterisation are especially concerning. They document the application of heated instruments to children's skin as treatment for symptoms including persistent diarrhoea, vomiting, cough, and fever. These observations bring a concrete child-safety concern into view. Respect for a family's beliefs must be accompanied by clear advice against harmful procedures and timely assessment of the underlying illness. Equally, the acceptance of a treatment by caregivers cannot establish its effectiveness or safety. The study records beliefs and experiences; these should not be interpreted as clinical validation of the remedies described.

Bamatraf recommends engagement with traditional healers because of their influence within the community. This proposal merits careful development around a defined purpose: recognising concerning symptoms, discouraging harmful practices, and

facilitating prompt referral. Such engagement should have explicit safety standards and should be evaluated. It must not imply professional endorsement of untested treatments. A useful programme would examine whether referrals occur earlier, whether families can act on them, and whether harmful practices become less frequent.

The paper also provides a basis for further research. Broader community studies could assess how frequently the reported beliefs and practices occur, including among families who do not use traditional medicine. Prospective work could examine the sequence and timing of treatment decisions. Educational interventions should be judged by changes in care-seeking and child safety, alongside knowledge. The value of this qualitative study lies in making those next questions more precise and locally relevant.

Khawla Faraaq Alzaitoni's review, Neutropenia in Infants, addresses a different diagnostic challenge. It describes conditions ranging from transient or self-limited neutropenia to serious inherited disorders. The practical emphasis is on interpreting laboratory findings together with age, clinical presentation, infection history, and changes over time. The review supports a proportionate approach that identifies infants needing specialist assessment while avoiding the assumption that every low neutrophil count carries the same implications. The other papers in this issue address related responsibilities in education, health-system organisation, and clinical interpretation. In *Invisible Curriculum*, *Unmeasured Outcomes*, Hakan Yaman and Aylin Yaman examine gaps between competency frameworks, assessment, accreditation, and feedback in family medicine training. Their conceptual analysis has a direct educational connection to Bamatraf's study: understanding a caregiver's explanation, discussing uncertainty, and negotiating a safe plan are abilities that require deliberate teaching and observation. A curriculum may name communication and community orientation as compe-

tencies without adequately establishing how trainees demonstrate them in practice.

Hakan Yaman's *Health System Expansion Without Proportional Health Gains?* examines the relationship between healthcare utilisation and outcomes in Türkiye. The paper raises questions about whether increasing service volume is accompanied by sufficient attention to appropriateness, continuity, and coordination. Its descriptive, aggregate analysis cannot establish that particular consultations or interventions were unnecessary. Nevertheless, it encourages scrutiny of what service activity achieves. Alongside the Yemeni study, it shows why both delayed access to needed care and potentially low-value utilisation require attention to the decisions and circumstances surrounding each encounter.

In *Cognitive Testing: Bedside Screening Tools and Formal Neuropsychological Assessment*, A. Abyad reviews the complementary roles of brief screening and detailed assessment. A central clinical message is that a score requires interpretation within the patient's history, functional abilities, language, education, and cultural context. Screening can identify the need for further evaluation, but it cannot supply every element of a diagnosis. This concern with interpretation is relevant across family medicine: measurement becomes useful when clinicians understand what it means for the individual patient.

The Letter to the Editor, *Data rich, meaning poor*, similarly argues for preserving the interpretation of patients' experiences within increasingly measurement-focused care.

Bamatraf's paper gives this issue its most immediate community-based priority. For practices caring for young children, its findings support a concrete change: make the family's understanding of illness, prior treatments, and ability to obtain further care a routine part of assessment. A safe plan should explain what to do at home, what to avoid, and when and where to seek help. The next step is

to test whether this approach enables children to receive appropriate treatment sooner.

Warm regards,
Dr. Abdulrazak Abyad
Editor-in-Chief
MEJFM

Chief Editor:

A. Abyad
MD, MPH, AGSF, AFCHSE
Email: aabyad@cyberia.net.lb
Mobile: 961-3-201901

Publisher

Lesley Pocock
medi+WORLD International
AUSTRALIA
Email:
publishermw@gmail.com

Copyright

While all efforts have been made to ensure the accuracy of the information in this journal, opinions expressed are those of the authors and do not necessarily reflect the views of The Publishers, Editor or the Editorial Board. The publishers, Editor and Editorial Board cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; or the views and opinions expressed.
p-ISSN: 1839-0188; e-ISSN : 1839-0196

<http://www.mejfm.com>

<http://www.worldfamilymedicine.com>

Editorial Board

http://www.mejfm.com/editorial_board.htm

Author Information:

http://www.mejfm.com/author_info.htm

Perceptions, Health-Seeking Behaviour and Traditional Treatment of Common Childhood Illnesses in Yemen: A Qualitative Study

Fauzia Faraj Bamatraf

MD, MSc, Professor of Community Health,
Department of Community Medicine College of Medicine and Health Sciences,
Hadhramout University (HUCOM),
Republic of Yemen

Correspondence:

Prof. Dr. Fauzia F Bamatraf
Department of Community Medicine, College of Medicine and Health Sciences,
Hadhramout University (HUCOM),
Al-Mukalla, Yemen
Mobile: +967 735306070

Email: drfauziabamatraf@gmail.com, ffbamatraf2008@yahoo.com

Received: August 2026. Accepted: September 2026; Published: October 2026.

Citation: Fauzia Faraj Bamatraf. Perceptions, Health-Seeking Behaviour and Traditional Treatment of Common Childhood Illnesses in Yemen: A Qualitative Study. World Family Medicine. October 2026; 24(7): 5-17 DOI: 10.5742/MEWFM.2026.241792

Abstract

Background: Although under-five mortality rates are declining globally, they remain high in developing countries, including Yemen. Deeply ingrained local perceptions and beliefs among caregivers regarding childhood illnesses are believed to influence how they diagnose and respond to these illnesses, contributing to high mortality rates.

Objectives: This study aimed to explore perceptions, health-seeking behaviours (HSB), and traditional medicine (TM) approaches for childhood illnesses in Al-Mukalla, Yemen.

Methods: A qualitative study was conducted that included three focus group discussion (FGDs) with caregivers, five semi-structured interviews (SSIs), as well as observations with traditional healers (TH). Participants were selected purposively. A total of 21 caregivers were included, each with at least one child under the age of five, who used TM or home remedies for their children during periods of illness, and who came from diverse socioeconomic backgrounds.

Results: Through FGDs and SSI, participants used the term “childhood illness” to describe any deviation from the ideal child state (“healthy child”) and considered it a normal part of child development. They also identified multiple causes for illness, combining medical and non-medical aspects. Some illnesses were categorized as “medically incurable” (such as those resulting from accidental falls, teething, or magic and the evil eye) and were believed to respond only to TM. This was the most common reason for delaying HSB. Common TM options in the study area included herbs, massage, cauterization, and Quranic healing.

Conclusion: The study underscores the need to raise awareness of caregivers in Al-Mukalla about the potential risks and complications associated with TM practices, particularly concerning the health of their children. TH are likely to play a role in treating childhood illnesses. Therefore, healthcare professionals should collaborate closely with them, given their significant influence. By providing them with accurate information, they can become agents of positive change in the community.

Keywords: Perceptions, HSB, TM, childhood illnesses, caregivers, TH, Yemen.

Introduction

Globally, under-five mortality rates (U5MR) have declined. However, child mortality rates are unacceptably high in developing countries, and the low uptake of healthcare for preventable childhood diseases is a major obstacle to increasing the use of available healthcare and reducing morbidity and mortality rates among children under five. In 2024, an estimated 4.9 million children under the age of five died worldwide (1). Approximately 75% of deaths among children under the age of five were caused by preventable and treatable diseases such as diarrhea, fever and acute respiratory infections (2,3).

Yemen is located in the southern part of the Arabian Peninsula and has a population of 39.4 million [4], making it one of the poorest countries in the Middle East [5]. In 2023, 40% of its population was under 15 years of age, and 16% was under 5 years of age [4]. Until 2013, Yemen had made steady progress in maternal and child health and in combating infectious diseases [6]. However, despite a gradual decline in the under-five mortality rate from 87 deaths per 1,000 live births in 1990 to 51 deaths per 1,000 live births in 2011, and projections of 38.1 deaths per 1,000 live births in 2024 [6,7], this rate is expected to rise again as a result of the ongoing conflict in Yemen since 2015, which has significantly exacerbated the country's chronic economic and health challenges. This has contributed to growing concerns about worsening poverty, leaving two-thirds of the population in need of humanitarian assistance [5].

The under-five mortality rate in Yemen is among the highest in the Middle East. Pneumonia was the most common cause of neonatal mortality, followed by diarrhea [8]. The health outcomes of children under five are significantly influenced by the behaviour of their healthcare providers, who are typically mothers, grandmothers, or other close family members [9]. Health seeking behavior (HSB) is defined as any action taken by individuals with health problems to obtain appropriate medical treatment for themselves or someone in their care [10].

In Yemen, caregivers' response to children under five is often delayed or inadequate. For example, national surveys show that only 35% of children seek care for conditions such as acute respiratory infections. Common obstacles include insufficient funding for treatment, distance to functioning health centers, and underestimation of the severity of illness by caregivers. Caregivers frequently delay treatment or resort to informal home remedies [11].

Delays in HSB are a contributing factor to high rates of child morbidity and mortality in developing countries, and providing appropriate medical care could prevent a significant proportion of child deaths and health problems [12]. Factors such as multiple ways to seek care and the inability to diagnose potentially life-threatening conditions have been found to be associated with these delays [13]. This delay can significantly impact a child's health, leading

to complications that reduce the effectiveness of medical care and, in some cases, render it ineffective. In his study in Al-Mukalla, Bamatraf reported that most children with persistent diarrhea were not referred for medical care until home remedies and Traditional Healers (TH) had failed [14]. During this period, the child may become severely dehydrated. Therefore, addressing the causes of these delays is crucial.

Cultural practices play a significant role in caregivers' behaviour when seeking healthcare and choosing treatment. The causes of illness are often categorized according to deeply held cultural beliefs, and healthcare decisions are consequently influenced by these beliefs. For example, while biomedical knowledge may help identify fever as a warning sign, cultural interpretations may attribute convulsions to witchcraft or the evil eye, leading caregivers to seek TH before resorting to formal healthcare facilities [15-17]. In developing countries, including Yemen, cultural and traditional practices for treating illnesses are widespread, with most caregivers preferring traditional remedies over modern healthcare services for treating their children's illnesses, largely due to their local availability [11,16,18, 19].

The World Health Organization (WHO) estimates that at least 80% of the population in most developing countries relies on traditional forms of primary healthcare [20]. Traditional Medicine (TM) continues to play a significant role in child healthcare in many Arab and developing countries [16,21-23]. In a study conducted by Bamatraf (2008) in Al-Mukalla, Yemen, it was observed that the majority of caregivers were more confident in home remedies they used themselves or in traditional healers to treat their children's diarrhea, due to a lack of trust in the services provided at health centers. They sought different treatment options before consulting any healthcare facility, leading to unnecessary delays in obtaining at least one consultation before hospital admission [14]. Ngere et al. [23] found an increasing demand for TM in western Kenya, where caregivers often prefer traditional methods for treating certain childhood illnesses. This study highlighted the need to improve safety and quality standards for traditional medicine products.

In general, there are published studies on caregiver behaviour patterns during childhood illness and the factors influencing health service utilization, particularly in developing countries. However, most of these studies are quantitative and often describe behaviour patterns without elucidating their causes and influencing factors, especially cultural practices and beliefs, which are widespread in developing countries [19,20,23]. In Yemen, cultural practices are widespread, with many believing that traditional medicine is always beneficial and never harmful [14,24]. Very few qualitative studies have been published, or are unavailable, on the factors influencing unhealthy caregiver behaviours in Yemen. A recent survey in Yemen revealed that 35% of children under five years old with respiratory symptoms and 53% with diarrhea did not receive medical care [11]. Therefore, this qualitative study aims to identify the perceptions, beliefs (regarding

study aims to identify the perceptions, beliefs (regarding the causes and severity of illness) and healthcare-seeking behaviours of caregivers when their children become ill. It also aims to identify the traditional treatment methods used by caregivers and traditional healers for these illnesses in Al-Mukalla, Yemen. Ultimately, this will lead to a better understanding of how to design effective policies to improve health behaviour and child health.

Methods and Materials

Study setting:

The study was carried out in Al-Mukalla, Hadhramout governorate, Yemen. It is located on the Hadhramout coast of the Arabian Sea in the South of Yemen. According to the last Complete Census Conducted in 2004, Al-Mukalla has about 196,187 inhabitants. Ongoing conflict has prevented a nationwide census in the last two decades [25]. In Al-Mukalla, health services are provided by the public sector, the private sector, and traditional healers. The public sector remains the primary provider of health care at all service levels [26].

Study participants and sampling procedures:

A qualitative study was conducted between January and May 2025, involving caregivers of children under five years of age. Purposive sampling method was used to select participants based on predefined inclusion criteria, including age, marital status, responsibility for the care of at least one child under five, and residence in Al-Mukalla. Caregivers were selected from diverse socioeconomic backgrounds and with prior experience using TM with their children, in order to encompass a wide range of perspectives on childhood illnesses and HSB. To enhance contextual understanding, five traditional healers (TH) were included in the study: two herbalists, a religious healer, a massager, and a cauterization healer. Their practices were observed. They were purposely selected based on their roles and involvement in the management of childhood illnesses.

Three focus group discussions (FGDs) were conducted with caregivers, each consisting of seven participants. Each FGD lasted between 60 and 90 minutes. The FGDs explored caregivers' perceptions, their initial reactions, HSB and treatment of childhood illnesses. The discussions were facilitated by the researcher and a trained assistant, with the assistant acting as facilitator while the researcher took notes. Responses were manually recorded after each session before proceeding to the next. In addition, five Semi- Structured Interviews (SSIs) were conducted with TH in their workplaces, either at home or in their shops. Each interview lasted approximately 45 minutes, allowing ample time for participants to discuss their perceptions and treatment experiences for common childhood illnesses in detail. Some of their treatment practices were also observed, to document their actual practices when bringing children in for treatment. Photographs were taken.

Data analysis

All SSIs and FGDs were audio-recorded with the participants' consent and transcribed verbatim. Content analysis was used to interpret the results and extract themes from the data. According to content analysis and the continuous comparison technique Pobe et al [27], the text was read repeatedly to familiarize participants with the content, then categorized into words and phrases to find meaningful relationships between emerging categories and themes. The described categories and themes were extracted inductively, that is, gradually and with the help of continuous re-reading and analysis.

Ethical approval

Ethical approval was obtained from the College of Medicine and Health Sciences, Hadhramout University (HUCOM). FGDs, SSIs and observation only commenced after obtaining informed consent and explaining the study objectives to the participants. The privacy and confidentiality of participant data were ensured. To maintain the confidentiality of the participants' identities and prevent their disclosure, the participants' identities were hidden (using numbers) in all documents, with these numbers being used throughout the study period.

Results

Thematic analysis of the data from FGDs, SSIs and observation were carried out. The results were presented within five interrelated themes: (1) participants' perceptions of the causes of childhood illness, (2) symptom severity and initial caregiver responses to illness, (3) caregiver decision-making processes (HSB), (4) treatment of childhood illness by caregivers and TH, and (5) observations on treatment practices by TH.

Theme 1 Participants' perceptions of causation:

In this study, participants in FGDs and SSIs used the term "childhood illness" to describe any deviation from the ideal child state ("healthy child"), but they did not consider it a serious illness, such as diarrhea, cough, fever, and vomiting, which are considered part of a child's development.

"These illnesses themselves don't kill" One grandmother said:

Most of the participants agreed with this sentiment.

Regarding the causes of childhood illnesses, participants in FGDs and SSIs cited numerous reasons and offered diverse explanations, combining both medical and cultural perspectives. They linked these illnesses to child development. One mother in a FGD said:

"This happens when the child starts teething, or when they start crawling or walking. It's not an illness; it's normal. It lasts for a few days and doesn't require any medical treatment."

Caregivers in FGDs also mentioned poor hygiene (a dirty bottle or hands, or improperly handling contaminated food). One caregiver said:

"If the child plays in dirty places and eats without washing their hands, they will get sick."

In addition, caregivers and healers attributed childhood illnesses, such as persistent diarrhea and vomiting in infants, as well as poor appetite, fever, and/or persistent cough, to accidental falls (Ragabah) when a child falls or is handled improperly. Caregivers believe that in this case the child does not need medical care but should be taken directly to a massage healer. One caregiver in FGD said,

"These illnesses are usually treated with traditional therapy (massage); medical treatment is not necessary".

A mother in another FGD said:

"My daughter fell once. We took her immediately to a clinic, but her condition did not improve for several days. I called my grandmother to give her a neck massage, and she started to get better that same day!"

Furthermore, most participants linked diarrhea and colic to unsuitable breast milk for their baby. One mother in FGD said,

"Because my milk isn't suitable, to prevent my babies from getting these illnesses, I started feeding them formula from birth, and it's beneficial."

Participants have linked worm infections to eating large amounts of sweets, while they have linked a child's high fever, difficulty breathing, or flu symptoms to exposure to cold weather or cold air. One mother in FGDs said,

"I think my baby gets sick most of the time because of the weather. When he's cold, he's more likely to get fever, difficulty of breathing or the flu."

Furthermore, the caregivers and healers attributed the illness to spiritual or supernatural causes (such as envy). Envy, they believed, occurs when people look at a child with malicious intent or when they wished that someone didn't possess what others had. Some caregivers disagreed with these views.

As one young caregiver put it, *"It doesn't make sense. We don't believe that anymore."*

In addition, they believed these illnesses were God's will. One grandmother said:

"It is God's will if a child has some of these illnesses."

All the caregivers agreed with this view.

Theme 2: Symptom severity and initial responses to childhood illness:

The recognition of a symptom is a critical factor that influences timely HSB among caregivers. In most childhood illnesses, the mother is the first to notice symptoms in her child. She then begins consulting her social support network, which usually includes older family members (grandmothers or mothers-in-law), and sometimes friends or neighbours with children, seeking advice on the nature of the illness and its treatment.

It was observed in the FGDs that out-of-home care and treatment were not sought immediately or within 24 hours of the child becoming ill. For most caregivers, the decision to seek out-of-home care often stemmed from a feeling that self-treatment was ineffective in alleviating the child's symptoms and that, in most cases, the child's condition was worsening. One mother in FGDs said:

"I suspected my five-year-old daughter was ill when she stayed home and her condition worsened; she developed a high fever, and I asked my mother-in-law for help in diagnosing it. We agreed it was a flu-like illness and that we would treat it with home remedies. Her fever suddenly spiked, and she began having difficulty breathing. That's when we realized it was a respiratory illness, and we needed to take her to the hospital."

As for the remaining caregivers, most explicitly stated their desire to seek out-of-home care immediately, especially when they felt the child's condition was serious.

One caregiver in FGDs said,

"There are no home remedies for high fever or difficulty breathing. If a mother keeps her sick child at home, they will die." All mothers should know this and act immediately.

Theme 3: Decision-making in healthcare seeking:

In this study, caregivers reported using multiple strategies and treatment options when their children were ill. Typically, caregivers begin treating the sick child at home with whatever is available to manage the illness. Home fever control, as reported in all FGDs, included applying cool compresses to the child's head and extremities, administering fever reducers such as paracetamol, and, in some cases, massaging the child with herbal oils. They also mentioned drinking honey, with or without herbs, or with black seed, as a treatment and preventative measure for childhood illnesses.

One caregiver in FGDs said:

"Yes, honey is beneficial for newborns, children, and adults. I have given my children honey since birth to maintain their health, and also as a treatment for diarrhea and other illnesses."

Rice water, herbal tea, and yoghurt were also mentioned as treatments for diarrhea. When diarrhea is accompanied by vomiting, some grandmothers massage the child's neck and/or abdomen, and sometimes they stop breastfeeding and restrict certain foods during the episode of diarrhea. In cases of worms, vinegar is given to the child to drink because it kills the worms.

One grandmother in FGDs said: *"Yes, drinking vinegar with a little herb cleanses the stomach of parasites."*

Participants also mentioned the use of over-the-counter medications (such as antibiotics, anti-diarrheal, and anti-emetics), leftover medications from previous treatments, and sometimes sharing medications among family members, as common practice. One caregiver said,

"Sometimes we give the child antibiotics early, as soon as they show symptoms."

Some participants in the FGDs nodded in agreement. Another caregiver added,

"I do the same thing with my children. If I go to the doctor, they'll prescribe the same medication."

When we asked the caregivers, “What medications do you give your sick children?” they couldn’t specify. However, a young caregiver explained,

“My husband usually buys it, and we always have it at home.”

Another caregiver in the same FGDs provided further clarification:

“I don’t remember the exact name, but it’s a pink liquid, and you can buy it at the pharmacy.”

The caregivers mentioned that, they sometimes use both types of services consecutively without waiting for results. As one mother explained,

“When my child is sick and I’m scared, I use whatever seems to work. When I have the money, I go to the hospital, and in the meantime, I use whatever my mother-in-law gives me.”

If the cause is believed to be envy or an accidental fall, they go directly to a traditional healer. One grandmother said, *“Some childhood illnesses don’t require a visit to a doctor or hospital.”*

Only a few caregivers mentioned using oral rehydration solution (ORS) to treat diarrhea at home. Help outside home would be sought only when the child did not get better, as one grandmother reported:

“We just start use traditional home remedies. If all fail, then we take the child to the health institution”.

Most of participants in other FGDs agreed with these sentiments. Home remedies were described as inexpensive and convenient by all participants in all FGDs

Theme 5: Treatment of Childhood illness practiced by traditional healers:

Herbalists provide treatments to residents of Al-Mukalla and the surrounding areas. Their clients include men, women, and their children from diverse socioeconomic backgrounds and age groups. Most seek consultations for themselves and their children suffering from various childhood illnesses. They receive patients in their homes or shops. The healers first listen to a brief explanation from the patient, then diagnose their condition based on questions and answers, and finally prescribe the appropriate herb or compound. Several plants have been mentioned as being used to treat childhood ailments. The healers prepare a powder from each plant, sometimes combining several herbs, and give it to the sick child. The recommended dose for infants is a pinch or half a teaspoon of this powder mixed with milk, boiled water, or honey, twice a day for up to three days. The dose and duration of treatment can be increased for older children. It is worth noting that these herbs were widely used for both adults and children, with variations in dosage or sometimes mixed with oil for external use. As one herbalist explained, *“Internal and external application are essential for a complete cure.”* They also reported that in rare cases, if symptoms persist or complications arise, special herbal formulations are recommended to treat the illness. Healers believe these herbs improve the child’s blood, address the root cause of the illness, and strengthen the body.

As for the spiritual or religious healer (*Alsheikh*), he treats childhood illnesses believed to be caused by envy or the evil eye through spiritual healing, reciting prayers and selected verses from the Quran. The Alsheikh writes these verses on a piece of paper placed under the child’s pillow or recites them over water or oil. The child then drinks the water over which the Quran has been recited, while the oil is applied to their body. Sometimes, in addition to the Quran, he advises consuming herbs or burning incense. Each session lasts ten minutes and continues for three days or more.

Theme 5: The observations:

The observation was conducted with two healers, a massager and a cauterization healer. They were purposely chosen as their methods (neck and/or abdominal massage and cauterization) are the commonest traditional practices available.

1. Cauterization (A harmful practice):

The healer is a woman in her sixties who has been practicing cauterization as a treatment for various symptoms and illnesses, exclusively for women and children, for over twenty years, having learned it from her mother. Her clients come from diverse socioeconomic backgrounds, including educated caregivers. This potentially harsh procedure is performed by healers, sometimes grandmothers, in the home. Most often, cauterization is done using a very hot iron rod, 10 to 30 mm in diameter, which is pressed firmly against the skin for 20 to 30 seconds until it turns white. The child's chest, abdomen and head were the most commonly used areas for cauterization. (Images 1,2) Thermal cauterization is used to treat chronic and intractable conditions, such as chronic or persistent diarrhea, flatulence, prolonged cough, persistent vomiting and excessive crying, and high fever accompanied by convulsions. It is even used to treat other ailments in children and adults, such as bone problems, polio, chronic wound cancers etc....



Image 1: Multiple cautery marks over a child's abdomen



Image 2: Multiple cauterization marks around a baby's navel

2. Massage treatment:

The healer was a 65-year-old woman with a primary education who had been practicing massage for 35 years. She inherited her massage therapy skills from her mother and grandmother, both of whom are deceased. Her clients were women and children brought in by caregivers. She saw her patients on the floor of a small, somewhat unclean room in her home. She described herself as a specialist in digestive and respiratory problems. She treated children of various ages, often suffering from symptoms resulting from accidental falls, such as diarrhea, vomiting, fever, persistent cough, loss of appetite, and general weakness. On the patient's first visit, the therapist would ask the caregiver some questions about the patient's medical history and whether they had experienced any accidental falls. The therapist observed the patient, focusing on their eyes, head, and limbs. She then massaged the child's neck with oil, moved their head up and down, and massaged their abdomen (Images 3, 4).

The massage session lasted about ten minutes, and the child cried and vomited at the end of the session. Caregivers typically recommended three to five more sessions of treatment. The treatment involved a massage with olive oil; the therapist used her bare hands and warm oil. The healer believes that massage opens the nerves and blood vessels in the neck and abdomen, improving blood flow to the stomach. Treatment typically lasts three to five sessions.



Image 3: A neck massage session for a 6-month-old child



Image 4: Abdominal massage session for a 7-month-old

In some cases, if symptoms persist, the healer recommends cupping without bloodletting. She asserts that cupping is the most effective and recommends continuing it for five days or more. The treatment tool is a dry cup. To perform the procedure, she first selects an area between the shoulder blades (4 cm below the last cervical vertebra), then places the dry cup on the area and begins suction for one minute. She repeats the suction two or three times in each session (Image 5).



Image 5: Cupping without bloodletting session for a 3-year-old girl

During the waiting period, the researcher conducted brief interviews with some of the caregivers about the reason for their visit and whether they had sought medical help before resorting to the healer. Most of the caregivers had tried home remedies without success, while some went directly to the TH, believing she was the only one who could help. However, none of them had sought medical help, and all were satisfied with the treatment results. According to participants in FGDs, this acceptance stems from the fact that:

“massage therapy is comfortable, practical, economical, and painless, and it does not cause any side effects in treating childhood diarrhea or other symptoms, so we are more willing to treat our children with it.”

Discussion

Our results indicate that the decline in access to health facilities for treatment is not due to a lack of awareness of it or even a lack of perceived need for this health option, but is related to social and cultural factors, such as how caregivers understand and perceive the child's illness, and how they deal with it within families and communities. The study shows that the participants presented a wide range of local perceptions and explanations, combining medical and cultural perspectives on childhood illnesses. These local concepts of illness and its causes significantly influenced the type of treatment used, whether caregivers would first resort to TM, use home remedies, or go directly to medical facilities. These explanations included natural and environmental factors, a normal stage of child development, poor hygiene, accidental falls, the evil eye, and God's will. These responses demonstrate the coexistence of medical knowledge and traditional beliefs about the concept and causes of illness. The ranges of perceptions of childhood illnesses among caregivers in FGDs largely matched those reported by TH in SSIs. There are also points of similarity with perceptions in the previous study conducted by the author (Bamatraf) in 2008, which reported that caregivers considered diarrhea a normal consequence of teething, expected with each tooth erupting, and not viewed as a serious illness requiring attention [14]. These findings are also consistent with other studies [28-30]. It is crucial that caregivers understand the meaning of the causes of childhood illness; only then can they make informed decisions and seek medical help. Numerous qualitative studies have emphasized the importance of considering cultural factors, such as local beliefs about the causes and classification of illnesses, because of their profound effects on the effectiveness and design of health promotion programs. [22,31,32]

The current study showed that the perceived severity of the illness was also one of the most common reasons for avoiding or delaying seeking medical care among caregivers. Although caregivers listed various causes of illness, some of which they considered not requiring medical attention, they associated high fever and difficulty breathing with severe illness requiring immediate medical

care. This aligns with the biomedical understanding of warning signs in conditions such as malaria or pneumonia [33]. In many resource-limited settings, delays in recognizing warning signs, such as fever, seizures, and difficulty breathing, lead to increased morbidity and mortality rates among children [33]. Identifying symptoms is often considered a major challenge in many studies addressing healthcare seeking for childhood illnesses [34]. Mothers and caregivers are often the first to notice symptoms of illness in their children. A lack of awareness of symptoms and warning signs by caregivers can lead to delays or failures in seeking treatment. These decisions are rarely simple; rather, they are intertwined within a complex web of cultural, religious, social, and economic perceptions and beliefs [35]. These findings are consistent with other studies showing that caregivers often adopt a pluralistic understanding of illness, combining natural and non-natural explanations. A caregiver's ability to differentiate between mild and severe illness often influences their decision regarding symptom management at home, seeking advice from TH, or resorting to modern medical healthcare facilities [14,29,36].

Despite the fact that the caregivers in the current study who participated in the focus group discussions had university degrees, secondary and primary qualifications, and belonged to diverse age groups, they all expressed a preference for TM, demonstrating the continued prevalence of traditional healing practices among individuals from diverse backgrounds. Our study showed that healthcare decision-making is influenced by a complex interplay of perceived causes, symptom severity, and cultural beliefs. Belief in TM was prevalent, even among those who primarily preferred modern medical facilities [36]. It is particularly important to note that caregivers identified multiple medical and non-medical causes for childhood illnesses. They demonstrated a high degree of similarity in their localized perceptions of the causes of these illnesses, and consequently, most sought care from multiple providers and pursued multiple treatment approaches simultaneously to achieve tangible benefit. TH were often perceived as more convenient, accessible, and less expensive. On the other hand, the severity of the illness was a strong motivator for seeking medical care, as the modern medical system was seen as a safe haven should the child's condition deteriorate. This was also a common finding in previous studies [16,19,22]. This classification of diseases shows that local beliefs about their causes and severity may hinder or delay the seeking of medical care. In a study conducted among caregivers in Al-Mukalla, Bamatraf found that many of the barriers to seeking care among children with diarrhea were related to local perceptions of the illness. Consequently, most children with persistent diarrhea were referred to medical care only after home remedies and traditional healers' treatments had failed. During this period, the child may become severely dehydrated [14]. These results are consistent with other studies that have shown the prevalence of medical pluralism and widespread belief in traditional remedies in many developing countries [16,22,36]. Furthermore,

our study identified several factors influencing HSB for childhood illnesses in both modern and traditional medical facilities. Belief in TM was prevalent, even among those who primarily preferred modern medical facilities. Consequently, TH were perceived as more suitable for treating these illnesses. The ranges of caregivers' perceptions of childhood illnesses in this study largely aligned with those reported by TH.

It is to be noted that the belief among caregivers that an illness does not require medical treatment was the most common reason for avoiding or delaying seeking medical care. It also became clear that when childhood illnesses are classified according to local beliefs as "uncurable by medicine" (such as those resulting from accidental falls, teething, magic, the evil eye, or fate), traditional healers are often seen as more appropriate, more accessible, and less expensive for these ailments.

As one caregiver stated,
"These illnesses are usually treated with traditional medicine; medical treatment is unnecessary."

This has significantly influenced HSB (ranging from home remedies and the use of over-the-counter medications to reliance on traditional healers). There is a widespread belief in Yemen and in many Arab and other developing countries that some diseases can only be "cured" or best prevented through TM, which contributes to delays in seeking healthcare services, diagnosis, and treatment [16,31,32,37,38].

Traditional medicine (TM) plays a crucial role in primary healthcare in many developing countries, and its use has increased in recent years. However, some elements of TM are beneficial, while others may not be. According to the WHO, more than 80% of the global population across over 170 countries uses traditional, complementary, and integrative medicine. People turn to these practices as a primary or secondary health resource [20].

The prevalence of TM use among caregivers and TH in treating childhood illnesses in Al-Mukalla was assessed, revealing that herbs was the most commonly used TM in the area. This finding aligns with the cultural and historical significance of herbal remedies and massage in traditional healing practices. However, despite the presence of herbal medicines in many countries, a relatively small number of plant species have been studied to determine their potential medical applications. Caregivers should be aware that the vast majority of herbal medicines have not undergone the same rigorous laboratory testing for safety and efficacy as Western medicines, nor have they been subject to the same level of government regulation.

Other common TM included oil ointments, honey, cupping, and Quranic healing. These results support the notion that traditional healing practices in the Al-Mukalla region encompass a wide range of approaches, reflecting a holistic approach to healthcare. TH and caregivers have long considered honey an important part

of traditional and religious treatments, both preventative and curative. Honey has been used for centuries as a natural remedy, and modern science confirms its powerful antimicrobial properties. It is effective against pathogens such as *Staphylococcus aureus* and *Candida albicans* [39]. Honey has also proven highly beneficial in treating infected surgical wounds [40].

Another TM was abdominal and neck massage, which was a common treatment for childhood symptoms (caused by accidental falls) and was practiced by TH and grandmothers in the study area. This treatment has been widely accepted by caregivers as a satisfactory remedy for their children. Professor Zhang from China noted this promising outcome of abdominal massage therapy. He reported that infants treated with traditional abdominal massage or by using ECIWO (Embryo Containing the Information of the Whole Organism) massage got faster recovery from diarrhoea, than those who were treated with Western drugs (antibiotics) [41]. Massage therapy has also been mentioned as a treatment for diarrhea in children in other previous studies [21,42].

Other common traditional remedies for childhood ailments, as reported by caregivers, include cupping and cauterization. Traditional Arab cauterization was known and practiced before the advent of Islam. The Prophet Muhammad (peace be upon him) did not favor cauterization, saying, "Healing is in three things: cupping, honey, and cauterization, but I forbid cauterization." In another narration, the Prophet (peace be upon him) said, "If there is healing in your medicine, it is in cupping, honey, or cauterization, but I do not like to be cauterized with fire" [43].

Cauterization was a common practice in Yemen and was mentioned in the National Survey as a treatment for umbilical cord injuries in newborns of Yemeni mothers [44]. It was also commonly used in many other Arab and Muslim countries, including, Saudi Arabia, Sudan, Oman, Libya, and Iran and, to treat various illnesses in children and adults [21,45-48]. Cauterization was also a common treatment in ancient Greek, Indian, and Chinese healing practices [49]. This aggressive procedure may have adverse consequences in infants and young children. Infection at the site of this procedure and death from sepsis may occur. Such practices should be depreciated.

Conclusion

The findings of this study confirm the persistence of traditional beliefs and practices in Al-Mukalla and their direct influence on how caregivers manage their children's illnesses. Misconceptions among caregivers regarding the causes of common childhood illnesses were reflected in case management and healthcare practices. In most cases, caregivers clearly placed more trust in home remedies or traditional healers than in the services offered at health centers. This tendency contributed to delays and limited use of healthcare services. Understanding

local beliefs is crucial for health development planning, and caregivers in Al-Mukalla should be educated about the potential risks and complications associated with traditional medicine practices, particularly concerning their children's health. TH are likely to play a role in the treatment of childhood illnesses, so healthcare professionals should collaborate closely with them given their significant influence. By providing them with accurate information, they can become agents of positive change in the community.

Acknowledgements

The author would like to express his sincere gratitude to all the caregivers and traditional healers who participated in the study for their generous cooperation.

References

- World Health Organization. Child mortality (under 5 children) 2025. Available at: <https://www.who.int/news-room/fact-sheets/detail/child-mortality-under-5-years> [Accessed May 2025]
- World Health Organization. World health statistics 2016: monitoring health for the SDGs sustainable development goals: WHO, 2016.
- World Health Organization. Levels and trends in child mortality. 2018.
- United Nations Department of Economic and Social Affairs Population Division. World Population Prospects 2024, Online Edition. 2024. Available at: <https://population.un.org/wpp/Download/Standard/Population>. [Accessed May 2025]
- World Bank Group. The World Bank in Yemen. <https://www.worldbank.org/en/country/yemen/overview>. 2024. Accessed 2024 July 16
- World Health Organization. Child and adolescent health, Yemen. 2024. <https://www.emro.who.int/child-adolescent-health/data-statistics/yemen.html>. Accessed July/ 2025
- UNICEF Data 2026: Monitoring the situation of children and women in Yemen Available at: <https://data.unicef.org/country/yem/>. Accessed February 2026
- Qirbi N, Ismail SA. Health system functionality in a low-income country in the midst of conflict: the case of Yemen. *Health Policy Plan*. 2017;32 (6): pp 911–22.
- UNICEF. Under-five Mortality 2025. Available at <https://data.unicef.org/topic/child-survival/under-five-mortality/>
- Olenja J. Editorial Health seeking behaviour in context. *East Afr Med J*. 2003;80 (2):61–62
- Central Statistical Organization (CSO) and United Nations Children's Fund (UNICEF). Yemen multiple Indicator Cluster Survey 2022–2023, Survey findings Report. Yemen: Central Statistical Organization and United Nations Children's Fund; 2023. <https://mics.unicef.org/news/just-releasedmics-yemen-survey-findings-report-and-snapshots>. Published 2023, [Accessed March/ 2025].
- Mpimbaza A, Nayiga S, Ndeezi G, et al.: Understanding the context of delays in seeking appropriate care for children with symptoms of severe malaria in Uganda. *PLoS One*. 2019; 14 (6): pp 62 - 72.
- Kolola T, Gezahegn T, Addisie M: Health care seeking behavior for common childhood illnesses in Jeldu District, Oromia Regional State, Ethiopia. *PLoS One*. 2016; 11 (10):2-8
- Bamatraf FF. Determinants, health care seeking behavior and treatment of persistent diarrhoea among children under five years of age in Al-Mukalla District. PhD thesis. University Kebangsaan Malaysia; 2008. (Unpublished data.)
- Muhanga M, Malungo JR. Does Attitude Associate, Correlate, or Cause Behaviour? An Assessment of Attitude towards Health Behaviour under One Health Approach in Morogoro, Tanzania. *Int J Adv Res Publications (IJARP)*. 2017;3(1); pp 82–91.
- Ngere S, Akelo V, Ondeng'e K, Ridzon R, Otieno P, Nyanjom M, Omoro R, Barr T, B. Traditional Medicine Beliefs and Practices among Caregivers of Children under Five Years—The Child Health and Mortality Prevention Surveillance (CHAMPS), Western Kenya: A qualitative study. *PLoS ONE*. 2022;17(11): e0276735. <https://doi.org/10.1371/journal.pone.0276735>.
- Mburu CM, Bukachi SA, Shilabukha K, et al.: Determinants of treatment-seeking behavior during self-reported febrile illness episodes using the socio-ecological model in Kilombero District, Tanzania. *BMC Public Health*. 2021; 21(1): pp1075-84
- Ajibade B, Amoo P, Adeleke M, Oyadiran G, Kolade O, & Olagunju R. Determinants of mothers health seeking behavior for their children in a Nigerian teaching hospital. *IOSR J Nurs Health Sci*; 2013, 1(6); pp 09–16
- Sultana M, Sarker AR, Sheikh N, Akram R, Ali N, Mahumud RA, Alam NH. Prevalence, determinants and health care-seeking behavior of childhood acute respiratory tract infections in Bangladesh. *PLoS One*. 2019. <https://doi.org/10.1371/journal.pone.0210433>.
- WHO. Integrating Traditional Medicine in Health Care. 2023. Available at: <https://www.who.int/southeastasia/news/feature-stories/detail/integrating-traditional-medicine>. [Accessed 12/2025]
- AlqahtaniYA, Shati AA, Alalyani RT, Alolah TA, Alshahrani MA, Asiri GB et al. Awareness and Attitude of Parents and Caregivers Regarding Harmful Traditional Medical Practices Towards Children in Aseer Region of Saudi Arabia. *International Journal of General Medicine* 2023;16; pp 4515–4524
- Al-Yateem N, Lajam AMA, Othman MMG, Ahmed MAA, Ibrahim S, Halimi A, Ahmad FR et al. The impact of cultural healthcare practices on Children's health in the United Arab Emirates: a qualitative study of traditional remedies and implications. *Front. Public Health*; 2023;11:1266742. doi: 10.3389/fpubh.2023.1266742
- Osarfo J, Ampofo GD and Tagbor HK. Health seeking behaviour of caregivers of children under five and its determinants in Ho West and Adaklu districts, Volta Region, Ghana: a community-based cross-sectional study. *BMC Public Health*; 2025 25:1219-26

24. Webair HH, & Bin-Gouth AS. Factors affecting health seeking behavior for common childhood illnesses in Yemen. *Patient Preference and Adherence*:2013; 7, pp 1129–38. <https://doi.org/10.2147/PPA.S51124>
25. Yemen Census. 2004. Central System Statistical in Hadhramout governorate, Republic of Yemen. December 2004
26. United States Agency for International Development (USAID). 2006. Yemen National Health Accounts: Estimate for 2003. Bethesda, MD: ABT Associates Incorporation.
27. Pope C., Ziebland S. & Mayas N. 2000. Qualitative research in health care: Analyzing qualitative data. *British Medical Journal* 320 (7272): pp114-116
28. Orora CN ,Obiri JF and Ombachi NK. Health Seeking Behaviour Determinants among Caregivers of Children Facing Diarrhoea in Homabay County, Kenya. *International Journal of Social Science and Humanities Research*. 2017; 5:3, pp 223-236
29. Msambal iIM, Myumbo L and Masanja P. Perceptions of illness and health-seeking behaviour among caregivers of under-five children in Morogoro municipality, Tanzania: a qualitative study *Discover Health Systems*. 2026; 5:50; pp 2-15
30. Ahorlu CK, Koram KA, Ahorlu C, De Savigny D, Weiss MG. Socio-cultural determinants of treatment delay for childhood malaria in southern Ghana. *Tropical Med Int Health*. 2006;11(7): pp1022–31.
31. Shaw B, Amouzou A, MillerNP, Bryce J and Surkan PJ. A qualitative exploration of care-seeking pathways for sick children in the rural Oromia region of Ethiopia. *BMC Health Services Research*. 2017; 17:184. DOI 10.1186/s12913-017-2123-5
32. Colvin CJ, Smith HJ, Swartz A, Ahs JW, De Heer J, Opiyo N, Kim JC, Marraccini T, George A. Understanding care seeking for child illness in sub-Saharan Africa: a systematic review and conceptual framework based on qualitative research of household recognition and response to child diarrhoea, pneumonia and malaria. *Soc Sci Med*. 2013;86: pp 66–78.
33. World Health Organization. 2024. Gender and health. Available at: https://www.who.int/he alth-topics /gender# tab=tab_1 [Accessed March/2026]
34. Riegel B, De Maria M, Barbaranelli C, Matarese M, Ausili D, Stromberg A, Jaarsma T. Symptom recognition as a mediator in the self-care of chronic illness. *Frontiers in public health*, 2022; 10, <https://doi.org/10.3389/fpubh>.
35. Merera AM. Determinants of acute respiratory infection among under-five children in rural Ethiopia. *BMC Infect Dis*. 2021. <https://doi.org/10.1186/s12879-021-06864-4>
36. Hooft A, Nabukalu D, Amumpaire JM, Gardiner MA, and Sundararajan R. Factors Motivating Traditional Healer versus Biomedical Facility Use for Treatment of Pediatric Febrile Illness: Results from a Qualitative Study in Southwestern Uganda. *Am. J. Trop. Med. Hyg*; 2020;103(1); pp 501–507
37. Tawfiq E, ID, Khwaja MIS, Alawi SAS, Jammalluddin J, Syed Nasir H. Predictors of Mothers' Care Seeking Behavior for Common Childhood Illnesses: Findings From the Afghanistan Health Survey 2015. *Int J Health Policy Manag* 2023;12, pp 7598-7609.
38. Simieneh MM, Mengistu MY, Gelagay AA, Gebeyehu MT. Mothers' health care seeking behavior and associated factors for common childhood illnesses, Northwest Ethiopia: community based cross-sectional study. *BMC Health Serv Res*. 2019;19(1):59. doi:10.1186/s12913-019-3897-4
39. Rana SH , Sawsan M A , Wjdan SQ , Ali MS. Antibacterial Effect of Natural Honey of known active components by GC-MS technique against *Staphylococcus aureus*. *Journal of Bioscience and Applied Research*, 2025,11: 2, pp 533-548.
40. Mohamed AEM. Benefits of using honey in the management of postoperative wound. *The Egyptian Journal of Hospital Medicine*. 2025;101, pp 5610-15.
41. Zhang X. 1996. Traditional medicine and WHO. *World Health 49th Year 2*.]
42. Abdul-Wahab, M.M. 1987. Traditional practices as a cause of morbidity and mortality in Juba area Sudan. *Annals of Tropical Pediatrics* 1: pp 18-21.
43. Deuraseh N. Health and medicine in the Islamic tradition based on the book of medicine (Kitab Al-Tibb) of Sahih Al-Bukhari. *JISHIM*. 2006; 5:2-14.
44. Yemen YDMCHS. 1997. Yemen Demographic and Maternal and Child Health Survey. Central Statistical Organization (CSO), Yemen and Macro International Incorporation. Calverton, Maryland: United States America. November 1998
45. Satti Abdelrahim, Mohamed-Omer SF, Hajabubker MA. Pattern and determinants of use of traditional treatments in children attending Gaafar Ibnauf Children's Hospital, Sudan. *Sudanese journal of paediatrics*. 2016;16(2): pp 45-50
46. Ghazanfar SA. Wasm: a traditional method of healing by cauterization. *Journal of ethno pharmacology*. 1995 Jul 28;47(3):pp 125-8.
47. Farid MK, El-Mansoury A. Kaiy (traditional cauterization) in Benghazi, Libya: complications versus effectiveness. *Pan African Medical Journal*. 2015; 22:98 doi:10.11604/pamj.2015.22.98.6399.
48. Asefzadeh, S., & Sameefar, F. 2001. Traditional healers in the Qazvin region of the Islamic Republic of Iran, a qualitative study. *Eastern Mediterranean Health Journal*, 7 (3), pp 544- 551.
49. Mohd Nayab, History of Amal-i-Kaiyy (Cauterization) and its indications according to the shapes of instruments: A review. *International Journal of Medical and Health Research*. 2017;3(3): pp 60-61.

Health System Expansion Without Proportional Health Gains? An Open-Data Policy Analysis of Healthcare Utilization and Overuse in Türkiye

Hakan Yaman, MD, MS, Professor

1. Tez Medikal Occupational Health Services, Antalya, Türkiye

2. American Hospital, Occupational Health Unit, Antalya, Türkiye

3. Independent Scholar, Antalya, Türkiye

4. Department of Family Medicine (Emeritus), Akdeniz University Faculty of Medicine, Antalya, Türkiye

Email: hakanyam@yahoo.com

Received: August 2026. Accepted: September 2026; Published: October 2026.

Citation: Hakan Yaman. Health System Expansion Without Proportional Health Gains? An Open-Data Policy Analysis of Healthcare Utilization and Overuse in Türkiye. World Family Medicine. October 2026; 24(7): 18 - 30 DOI: 10.5742/MEWFM.2026.241793

Abstract

Background: Health system reforms in Türkiye have substantially expanded healthcare access and utilization. However, whether increased service use has translated into proportional improvements in population health outcomes remains unclear.

Methods: This study employs a descriptive policy analysis combined with secondary analysis of open-access data from OECD, and national sources. Trends in healthcare utilization, selected population health outcomes, and system performance between 2003 and 2023 were examined descriptively using open-access OECD, World Bank, and national data, supplemented by comparative benchmarking and a narrative utilization–outcome gap framework.

Results: Healthcare utilization in Türkiye increased markedly, with physician visits per capita rising above OECD averages. Patient satisfaction followed a similar pattern, showing early gains followed by stabilization or decline. These findings indicate a growing divergence between service volume and health outcomes. Additional signals, including hospital-centered utilization patterns and evidence relevant to potentially avoidable hospitalizations, suggest emerging inefficiencies and potential overuse.

Conclusion: Türkiye appears to be transitioning from an access-expansion phase to a system maturity phase characterized by high utilization and a descriptive pattern compatible with diminishing marginal returns. Future improvements in population health will depend less on increasing service volume and more on enhancing care quality, coordination, and appropriateness. Policy priorities should include strengthening primary care gatekeeping, reducing low-value care, aligning incentives with outcomes, and improving data transparency. The Turkish case provides broader lessons for health systems globally, highlighting the need to shift from utilization-driven to value-based models of care.

Keywords: Health transformation, healthcare utilization, overuse, low-value care, Türkiye, health policy, OEC

Introduction

Health system reforms in Türkiye since 2003 have been widely recognized for expanding access and achieving near-universal coverage. These reforms significantly increased healthcare utilization through improved financial protection and service availability. However, the relationship between utilization and population health outcomes is not linear. Evidence from OECD countries shows that while early increases in healthcare use are associated with improved outcomes, these gains tend to diminish as systems mature and utilization rises [1,2].

As health systems mature, additional healthcare inputs may generate progressively smaller incremental health gains [1]. In some systems, excessive care is associated with inefficiencies and patient harm, reflecting a shift toward **overuse** [3,4].

Recent multi-country analyses show that while healthcare spending is associated with reduced mortality, broader outcomes such as quality of life and disability-adjusted measures depend more strongly on social determinants and non-medical factors [5,6]. This suggests that expanding healthcare utilization alone is insufficient to sustain long-term population health gains.

Against this background, a critical policy question emerges:

Does increased healthcare utilization necessarily lead to improved population health, or does it reflect a transition toward inefficiency and low-value care?

This paper explores the hypothesis that increased service utilization may not proportionally improve health outcomes and may instead contribute to overuse and inefficiency.

This study examines the hypothesis that rising healthcare utilization does not necessarily translate into proportional health gains and may contribute to systemic inefficiencies and low-value care. While expanded access can generate early improvements, excessive use of diagnostics and specialist services may lead to **overdiagnosis, overtreatment, and avoidable costs without clinical benefit** [7-9].

Overuse is increasingly understood as a **system-level phenomenon**, driven by weak gatekeeping, volume-based incentives, and high patient demand rather than isolated clinical decisions [10]. Consequently, increasing service volume may not improve outcomes and can instead expose patients to unnecessary risks while straining healthcare resources [11].

Conceptual framework

1 Healthcare Utilization vs. Health Outcomes

Healthcare systems aim to improve population health through accessible and effective care. In early phases, increased utilization, such as higher consultation and admission rates, is associated with substantial gains in mortality reduction and life expectancy, largely driven by improved access and management of acute and preventive care [5].

However, this relationship is not linear. Additional utilization yields **diminishing marginal returns**, with limited impact on population health in high-use settings [1]. This reflects a shift from addressing unmet need to potential **overuse of medical services**.

Excessive utilization may also cause harm. Overdiagnosis, unnecessary procedures, and excessive treatments expose patients to avoidable risks, while contributing to **resource misallocation** and diverting attention from high-value and preventive care [3,12].

Overall, improving health outcomes depends not on increasing care volume, but on ensuring the **appropriateness, quality, and value** of services delivered.

2 Overuse and Low-Value Care

Overuse, commonly termed **low-value care**, has become a central concern in contemporary health policy. It refers to services with little or no clinical benefit, such as excessive imaging, redundant testing, inappropriate pharmacotherapy, and avoidable procedures, which may increase costs and expose patients to harm [3].

International evidence indicates that overuse is widespread across health systems, with an estimated **10–30% of healthcare spending** attributable to wasteful care [11]. These patterns reflect systemic features rather than isolated clinical decisions.

Several structural drivers sustain overuse. **Fee-for-service payment models** incentivize higher service volume [4], while **defensive medicine** encourages unnecessary testing to mitigate legal risk [12]. In addition, **weak gatekeeping mechanisms** and direct access to specialist care promote fragmented and excessive utilization [13]. **Patient expectations**, often equating more care with better care, further reinforce these dynamics [10].

Beyond inefficiency, overuse represents a **patient safety risk**, contributing to overdiagnosis, overtreatment, and iatrogenic harm. Addressing low-value care therefore requires coordinated policy responses, including payment reform, strengthened primary care, evidence-based guidelines, and improved patient engagement.

Methods

1 Study Design

This study employs a **descriptive policy analysis combined with secondary data analysis** to examine the relationship between healthcare utilization and population health outcomes in Türkiye.

The analytical framework draws on health system performance literature, particularly the distinction between inputs, outputs, and outcomes, and the shift toward **value-based care** [2,14].

Rather than establishing causal relationships, the analysis focuses on identifying **macro-level trends, temporal patterns, and potential inefficiencies**, particularly the alignment between healthcare utilization and health outcomes over time.

2 Data Sources

The analysis integrates multiple **open-access and internationally comparable datasets** to ensure robustness through data triangulation. Healthcare utilization and system indicators were obtained from OECD Health Statistics and *Health at a Glance* reports, which provide standardized cross-country measures [2]. Population health outcomes were obtained from the World Bank World Development Indicators (WDI), including life expectancy at birth (SP.DYN.LE00.IN) and infant mortality (SP.DYN.IMRT.IN) [15,16].

Country-specific healthcare utilization data for Türkiye were obtained from the Ministry of Health Health Statistics Yearbook [17]. Additional contextual insights were drawn from peer-reviewed studies on the Turkish Health Transformation Program and its impact on access, equity, and system performance [18].

Patient satisfaction with health services was obtained from the Turkish Statistical Institute (TÜİK) Life Satisfaction Survey, which has been conducted annually since 2003. Changes in the handling of the “no opinion” response category from 2014 onward were considered when interpreting long-term satisfaction trends [19,20].

The integration of these sources enables **triangulation of evidence**, enhancing validity and reducing reliance on any single dataset.

3 Study Period

The analysis focuses on the period 2003–2023, corresponding to the implementation and maturation of Türkiye’s Health Transformation Program. This period captures three distinct phases:

The initial reform and expansion phase (2003–2012)

The consolidation phase (2013–2019)

The recent period of system pressure and recalibration (2020–2023)

This temporal segmentation enables the identification of structural shifts in both utilization and health outcomes.

4 Indicators and Operational Definitions

A set of indicators was selected to capture multiple dimensions of health system performance, consistent with OECD and WHO frameworks [2,21,22].

Healthcare utilization was assessed using total physician visits per capita and physician contacts at primary versus secondary/tertiary levels of care, reflecting service intensity and the distribution of healthcare contacts across care levels [2,23].

Health outcomes included life expectancy at birth and infant mortality, representing complementary population-level indicators of long-term health improvement [15,16,24].

System performance was evaluated through patient satisfaction and access indicators, capturing perceived quality and responsiveness [21,22].

Given the limitations of directly measuring efficiency at the macro level, **proxy indicators** were employed, including divergence between utilization and outcomes, avoidable hospitalizations, and patterns of low-value care. These proxies are widely used in OECD frameworks to identify inefficiencies in health systems [11].

5 Analytical Strategy

The analysis followed a **three-step analytical strategy**.

First, **descriptive trend analysis** was used to examine time-series changes in healthcare utilization and health outcomes, identifying growth patterns and temporal dynamics.

Second, **comparative benchmarking** positioned Türkiye’s performance relative to OECD averages, providing an international context for utilization and outcome indicators [2,11].

Third, a **utilization–outcome gap analysis** assessed whether increases in healthcare use corresponded to proportional improvements in outcomes. Periods in which healthcare utilization increased without commensurate improvement in selected outcomes were interpreted as descriptive signals of declining marginal value and possible system inefficiency. Such divergence was not considered direct evidence of overuse, which requires service-specific assessment of clinical appropriateness [14].

6 Conceptual Assumptions

The interpretation was informed by the concept of diminishing marginal returns, whereby additional healthcare inputs may produce progressively smaller health gains as unmet need is reduced [5]. Second, not all healthcare services contribute equally to improved outcomes, reflecting variation in **clinical value** [11]. Third, high utilization may simultaneously indicate improved access and **systemic overuse** [3].

7 Ethical Considerations

This study is based exclusively on publicly available, aggregated datasets. No individual-level or identifiable data were used. Therefore, ethical approval was not required.

8 Limitations

Several limitations should be acknowledged. First, the descriptive design does not permit causal inference. Second, aggregate data may obscure regional inequalities and population-level heterogeneity. Third, efficiency is assessed indirectly through proxy indicators rather than direct measurement. Finally, the influence of social determinants of health cannot be fully isolated within the scope of this analysis.

The observed increases represent descriptive trends derived from aggregated data and were not subjected to formal statistical hypothesis testing.

Results

1 Trends in Healthcare Utilization

Descriptive trend analysis shows a substantial and sustained increase in healthcare utilization in Türkiye following the Health Transformation Program. Physician visits per capita increased from 3.4 in 2003 to 11.4 in 2023. The long-term rise was temporarily interrupted during the COVID-19 period, with visits declining from 9.8 in 2019 to 7.2 in 2020, before recovering to 8.0 in 2021, 10.0 in 2022, and 11.4 in 2023 [2,23]. This long-term increase represents one of the clearest consequences of the expansion of healthcare access following the Health Transformation Program. Although utilization fluctuated during the COVID-19 period, the broader trajectory remained one of markedly increased contact with healthcare services.

Similarly, physician contacts at secondary and tertiary levels of care also increased substantially, from 2.1 per capita in 2003 to 6.5 in 2023, reflecting a marked increase in the use of secondary- and tertiary-level services [23]. This increase reflects expanded hospital capacity, improved financial protection, and reduced out-of-pocket barriers under universal coverage [2].

The rapid growth in hospital capacity, particularly following large-scale investments in public hospitals and city hospital projects, has facilitated greater reliance on secondary and tertiary care services.

Hospital-based service use also expanded during the reform period, reflecting increased capacity, broader financial protection, and easier access to secondary and tertiary care. However, high hospital utilization may represent both improved access and persistent reliance on hospital-centered pathways, particularly where primary care gatekeeping remains limited [14]. Such patterns are typical of systems with **weak gatekeeping and volume-based incentives**, which prioritize service intensity over appropriateness [11].

International evidence indicates that reliance on hospital-based care for conditions manageable in outpatient settings may signal **inefficiency and unnecessary utilization**. Accordingly, rising hospital use should be interpreted not only as improved access but also as a potential indicator of system-level **overuse and weak care coordination** [22,25].

2 Comparative Benchmarking with OECD Countries

Comparative benchmarking indicates that Türkiye exhibits a distinctive pattern of **high utilization relative to moderate health expenditure**. In 2021, Türkiye recorded 8.0 physician consultations per capita compared with an OECD average of 6.0 [2]. This diverges from the typical OECD pattern, where higher utilization is associated with higher expenditure.

This configuration suggests a **high-contact, cost-contained model**, raising concerns about consultation quality and clinical appropriateness. High visit volumes may constrain consultation time and reduce the effectiveness of clinical encounters [26,27]. From a systems perspective, this pattern reflects both **supply- and demand-driven pressures**, rather than purely need-based utilization. Weak gatekeeping and easy access to higher levels of care may increase service intensity without improving coordination or effectiveness. Health outcomes have improved substantially, including gains in life expectancy and reductions in maternal and infant mortality, largely driven by expanded access to essential services [2,28].

This pattern reflects **diminishing marginal returns**, whereby early expansion yields substantial gains, while further increases in utilization generate limited improvements [5]. Türkiye thus appears to be transitioning from addressing unmet need to a phase where **efficiency, appropriateness, and value of care** are the primary determinants of further progress. Taken together, these trends suggest a transition in the policy challenge facing the Turkish health system. During the earlier reform period, expanding access and utilization addressed substantial unmet need. As coverage matured and service use increased, however, further progress became increasingly dependent on the quality, appropriateness, coordination, and value of care rather than on service volume alone.

3 Trends in Health Outcomes

Life expectancy increased substantially over the study period, from 72.9 years in 2003 to 77.7 years in 2019. The COVID-19 period was associated with a temporary decline, followed by recovery to 77.2 years in 2023 (29). Life expectancy increased markedly during the initial phase of the Health Transformation Program, reflecting gains in access, preventive care, and maternal-child health services [2,28]. These early gains are widely attributed to the reduction of unmet healthcare needs and improved coverage.

However, improvements in life expectancy became more incremental, indicating a shift from rapid catch-up to slower progress. Infant mortality declined substantially, from 24.4 deaths per 1,000 live births in 2003 to 12.8 in 2012 and 9.4 in 2019. The absolute reduction was larger during the earlier reform period than during subsequent years (24). This suggests that initial gains were driven by expanded basic and preventive services, while further improvements require **system-level optimization** [5].

From a comparative perspective, this pattern is consistent with **diminishing marginal returns**, particularly in systems with broad coverage. Once essential service gaps are addressed, further increases in utilization yield **smaller incremental health gains** [30,31].

The stabilization of health outcomes despite rising utilization indicates a **utilization–outcome decoupling**, suggesting limited marginal effectiveness of additional inputs. Future health gains are therefore more likely to depend on improvements in **care quality, coordination, and value**, rather than further increases in service volume [21,30,32]. Thus, the long-term pattern is compatible with progressively smaller incremental health gains as the system moved from rapid expansion toward a more mature phase. This descriptive pattern should not be interpreted as evidence that increasing utilization itself caused the slower improvement in population health.

4 Patient Satisfaction and System Responsiveness

Patient satisfaction and perceived system responsiveness in Türkiye show a pattern similar to utilization trends, with **early improvement followed by stabilization or partial decline**. Satisfaction with health services increased from 39.5% in 2003 to 75.9% in 2011. It subsequently fluctuated and was 67.1% in 2019, 65.6% in 2022, and 65.4% in 2023. These data indicate substantial early improvement followed by lower levels relative to the earlier peak (19). Interpretation of the long-term series should account for the change in handling of the “no opinion” category introduced in 2014 and incorporated into published calculations from 2015 onward (20). During the initial phase of the Health Transformation Program, satisfaction increased markedly, driven by expanded coverage, improved access, and reduced waiting times [2,18].

However, more recent evidence suggests that this upward trend has plateaued, with some indicators pointing to declining satisfaction in specific domains of care experience. Emerging concerns increasingly relate to **shortened consultation times, perceived reductions in care quality, and persistent difficulties in accessing physicians despite high visit rates**. This apparent contradiction reflects a structural tension within high-utilization systems.

From a systems perspective, the coexistence of high utilization and declining perceived quality reflects **capacity strain and workflow fragmentation**. Increased consultation volumes reduce time per encounter, limiting clinical assessment, communication, and continuity of

care. Short consultation times have been associated at the health-system level with several adverse organizational and health-service indicators, including higher rates of ambulatory-care-sensitive hospital admissions [26,27,33].

High-frequency service use does not necessarily improve patient experience when care pathways are **fragmented and poorly coordinated**. In systems with weak gatekeeping, repeated and uncoordinated contacts can lead to inefficiencies, duplication, and patient dissatisfaction [11].

Taken together, these findings reinforce the notion that **increasing healthcare utilization alone is insufficient to improve system responsiveness**. Instead, patient satisfaction appears to be increasingly shaped by qualitative dimensions of care, including continuity, communication, and coordination. This underscores the need for a shift from volume-based performance metrics toward **patient-centered and value-based care models**, particularly in health systems that have already achieved high levels of service utilization.

The coexistence of high utilization with less favorable trends in some dimensions of patient experience suggests that service volume alone is an incomplete measure of health-system performance. Patient experience may increasingly depend on consultation quality, continuity, communication, and coordination.

5 Utilization–Outcome Gap Analysis

The analysis reveals a growing **decoupling between healthcare utilization and health outcomes** in Türkiye. While physician visits increased markedly, improvements in selected population-level outcomes became more incremental over time. This descriptive divergence is compatible with diminishing marginal returns but does not establish a statistical threshold, change point, or causal relationship [30,31].

This divergence is marking a shift from rapid gains to relative stabilization in outcomes.

- **Physician utilization increased markedly and exceeded OECD averages in recent comparable years.**
- **Life expectancy and infant mortality improved substantially over the long term, although subsequent gains were more incremental and were disrupted during the COVID-19 period.**
- **Satisfaction with health services increased markedly during the early reform period but was below its earlier peak in recent years.**

Together, these trends indicate a widening **utilization–outcome gap**, where increased service volume is no longer matched by proportional gains in health or patient experience. This pattern is consistent with **diminishing marginal returns**, whereby initial expansion yields

substantial gains, but further utilization produces limited benefits [5,11]. As healthcare utilization continued to rise, additional increases were accompanied by progressively smaller improvements in selected population-level outcomes. This descriptive divergence is compatible with diminishing marginal returns, although it does not identify a specific utilization level or establish a causal relationship. The emerging utilization–outcome gap therefore raises questions regarding the marginal value, appropriateness, and organization of additional healthcare activity.

The coexistence of rising utilization, plateauing outcomes, and declining satisfaction indicates a **structural, not merely quantitative, problem**. This gap reflects a misalignment between service volume and care value, driven by weak gatekeeping, fragmented care pathways, and volume-based incentives [11,34].

Comparative OECD evidence supports this interpretation: systems with **lower utilization but stronger primary care coordination** achieve comparable or better outcomes and higher patient satisfaction [11]. In contrast, high-contact systems without effective integration may generate **overuse without performance gains**.

Overall, the findings suggest that Türkiye's health system is entering a phase in which further improvements in population health will depend less on increasing utilization and more on enhancing the appropriateness, coordination, and value of care delivery.

6 Avoidable Hospitalizations and Signals of Overuse

The findings of this study provide converging signals of potential system-level inefficiency and overuse, although overuse was not directly measured at the level of individual services. Multiple indicators point toward a structural shift from a phase of unmet need reduction to one characterized by **potential overutilization and low-value care delivery**.

First, physician consultation rates in Türkiye exceed OECD averages, indicating a **high-contact model** that may not be fully aligned with clinical necessity. While initially reflecting improved access, persistently high utilization raises concerns about **appropriateness and efficiency** [2].

Taken together, these findings indicate that Türkiye's health system is undergoing a transition from an **access-expansion paradigm to a maturity phase** in which the dominant policy challenge is no longer increasing service volume, but rather ensuring that care delivery is **appropriate, coordinated, and value-based**. Without such a shift, continued growth in utilization risks reinforcing inefficiencies and undermining the long-term sustainability of the system. Comprehensive longitudinal national data on **ambulatory care sensitive conditions** (ACSC) admissions remain limited in Türkiye. Avoidable hospitalization is therefore used here primarily as a policy-relevant conceptual indicator, supported by International evidence, which generally suggests that better access to effective primary care is associated with lower rates of avoidable hospitalization, although findings across studies are heterogeneous.

6.1 Conceptual Framing

Avoidable hospitalizations, defined as admissions for **ACSCs**, are a key indicator of health system performance, particularly reflecting primary care effectiveness and care coordination [11,35]. These include conditions such as asthma, COPD, diabetes complications, and hypertension, which are largely manageable through timely outpatient care.

From a health systems perspective, avoidable hospitalizations act as a **proxy for inefficiency**, reflecting failures in prevention, early diagnosis, and continuity of care. Elevated ACSC admissions typically signal **weak primary care capacity, inadequate coordination, inappropriate access pathways, and overreliance on hospital-based care**.

Importantly, these admissions are shaped not only by clinical need but also by **system design, provider incentives, and patient behavior**. In systems with weak gatekeeping and fragmented care pathways, repeated and uncoordinated care increases the likelihood of avoidable admissions [11,13,36].

Comparative OECD evidence shows that **lower avoidable hospitalization rates** are associated with strong primary care systems characterized by effective gatekeeping and integrated care, whereas higher rates reflect misalignment between utilization and care value [11,13,35,36].

In this context, avoidable hospitalizations provide a critical lens for understanding the **utilization–outcome gap**, linking increased service volume to deficiencies in care quality and system organization.

6.2 Descriptive Findings

International comparative evidence generally suggests, although findings are heterogeneous that health systems with **strong primary care orientation and effective gatekeeping mechanisms** tend to exhibit lower rates of avoidable hospital admissions, particularly for ACSCs. In contrast, systems characterized by **high specialist accessibility, weak referral systems, and fragmented care pathways** are associated with higher rates of potentially avoidable hospitalizations [11,35].

In the Turkish context, although comprehensive longitudinal datasets specifically measuring avoidable hospitalizations remain limited, **available proxy indicators and indirect evidence suggest patterns consistent with increased hospital dependence and potential inefficiency**.

First, as demonstrated in Section 4.2, **overall hospital utilization rates have increased substantially**, reflecting expanded access but also indicating a persistent reliance on secondary and tertiary care services. Second, despite a marked increase in **physician contacts per capita**, there is no clear evidence of a proportional reduction in hospital admissions, which would be expected in systems where primary care effectively manages chronic and preventable conditions. This mismatch points toward potential inefficiencies in care delivery and coordination.

Third, the continued use of hospital-based services for conditions that are typically manageable within primary care settings suggests a **functional gap in primary care capacity and problem-solving ability**. This pattern may be further reinforced by weak gatekeeping, allowing patients to bypass primary care and directly access higher levels of care, thereby increasing system costs without commensurate improvements in outcomes [36].

Taken together, these findings indicate that **increased access to healthcare services has not been fully translated into effective primary care management**. As a result, the expected reduction in avoidable hospitalizations, commonly observed in well-integrated primary care systems, may be attenuated. This reinforces the broader conclusion that **service expansion without parallel strengthening of primary care functions and care coordination mechanisms may limit efficiency gains and contribute to overuse**.

6.3 Interpretation within Utilization–Outcome Gap

When interpreted alongside the **utilization–outcome gap**, avoidable hospitalizations provide an additional explanatory layer. Despite increased physician visits and expanded infrastructure, there is limited direct evidence; however, **available proxy indicators and indirect evidence suggest** a disconnect between service volume and care effectiveness.

This divergence suggests that utilization is not optimally distributed across care levels but instead concentrated in **higher-cost, hospital-based services**. Such patterns are consistent with weak gatekeeping, fragmented care pathways, and misaligned incentives, which constrain the coordinating role of primary care [11,13,35].

From a systems perspective, avoidable hospitalizations provide a plausible mechanism linking weaknesses in primary care and care coordination with high hospital utilization and suboptimal system performance.

This interpretation is consistent with international evidence showing that higher utilization does not inherently improve outcomes in the absence of **strong primary care integration and coordinated care delivery** [14]. Consequently, the Turkish experience illustrates how expanding access without simultaneously strengthening the **functional capacity and coordinating role of primary care** may result in inefficiencies and reduced system performance.

6.4 Link to Overuse and System Incentives

From a policy perspective, avoidable hospitalizations reflect not only **primary care underperformance but also system-level overuse driven by structural incentives**. Admissions for ACSCs may therefore indicate the combined effects of institutional arrangements shaping care-seeking behavior and provider decisions.

A key driver is **weak gatekeeping**, which allows direct access to hospital services and increases admissions for conditions manageable in primary care. This undermines care coordination and reinforces a **hospital-centered model of delivery** [13,37].

Second, **performance-based payment systems** that reward service volume may incentivize excessive diagnostic and therapeutic activity. Such incentives create pressure to increase throughput rather than optimize appropriateness, contributing to the expansion of **low-value care** [4,11].

Third, **fragmented care pathways** further exacerbate inefficiencies. In the absence of integrated care models, patients may experience repeated, uncoordinated encounters across different providers and levels of care, increasing the risk of duplication, unnecessary admissions, and discontinuity in disease management [36].

Fourth, **patient behavior and expectations** also play a role. In systems where hospital services are perceived as more accessible, comprehensive, or higher quality, patients may preferentially seek hospital-based care, reinforcing demand for specialist and inpatient services even when primary care would be clinically appropriate.

Taken together, these factors suggest that avoidable hospitalizations in Türkiye may function as a **substitute for insufficiently integrated primary care**, rather than as a medically necessary endpoint of care. This interpretation reinforces the broader conclusion that overuse is not solely a clinical issue but a **systemic phenomenon shaped by incentives, access pathways, and institutional design**.

Addressing this challenge therefore requires policy interventions that realign incentives, strengthen gatekeeping, and enhance the **coordinating and problem-solving capacity of primary care**, rather than simply reducing service volume.

6.5 Policy-Relevant Implications

The potential burden of avoidable hospitalizations highlights the importance of examining the alignment between access, utilization, and care allocation in Türkiye. Although the Health Transformation Program expanded coverage and service use, **value-oriented and efficient care delivery remains suboptimal**.

This misalignment can be summarized as follows:

First, strengthening **primary care gatekeeping** is essential to ensure appropriate first-contact management and reduce unnecessary referrals. Evidence shows that strong primary care systems are associated with lower avoidable hospitalizations and greater efficiency [35,37].

Second, implementing **integrated care coordination**, particularly for chronic disease management, is essential. Effective coordination across care levels reduces duplication, improves continuity, and prevents avoidable hospital admissions [35].

Third, routine monitoring of **ACSC admission rates** should be institutionalized as a key performance indicator. These metrics provide a practical and internationally comparable tool for evaluating primary care effectiveness and identifying system inefficiencies [2,11].

Fourth, **realigning incentives toward value-based care** is essential to reduce overuse. Volume-driven payment models may encourage unnecessary interventions, whereas value-based approaches promote **appropriate, patient-centered care**.

Finally, patient behavior and expectations should be addressed through health literacy initiatives and system navigation strategies, encouraging appropriate use of primary care services and reducing unnecessary reliance on hospital-based care.

These mechanisms can be understood as a reinforcing policy pathway. Weak gatekeeping and limited primary care coordination may encourage hospital-centered utilization; greater hospital dependence may in turn reinforce fragmented care pathways and volume-oriented incentives. Within the Walt and Gilson framework, these relationships reflect interactions between policy context, actors, processes, and content rather than a single causal mechanism. The model is framed within the Walt and Gilson policy triangle, which emphasizes the interaction of policy context, actors, processes, and content in shaping health system performance [38].

7 Summary of Key Findings

The findings indicate a **structural transition** in the Turkish health system, characterized by high utilization, slowing outcome gains, and emerging inefficiencies.

First, healthcare utilization has increased substantially, with physician visit rates exceeding OECD averages, reflecting successful expansion of access. Second, health outcomes have improved, but gains have slowed over time, indicating a shift from rapid early gains toward more incremental improvements in selected population-health outcomes. Third, a growing **utilization–outcome gap** has emerged, where rising service volume is no longer matched by proportional improvements in population health or patient satisfaction, supported by patterns consistent with avoidable hospitalizations and perceived declines in care quality.

Together, these findings suggest that the central policy challenge has shifted from expanding access to ensuring **efficient, well-coordinated, and value-based care delivery**. Without this transition, further increases in utilization risk reinforcing inefficiencies and overuse rather than improving health outcomes.

Discussion

1 The Utilization–Outcome Paradox

The Turkish experience illustrates a fundamental health system paradox: increased utilization does not necessarily translate into improved population health outcomes. While expanded access has produced substantial early gains, as systems mature, additional inputs may yield progressively smaller gains [1,5].

Several interrelated mechanisms may explain this utilization–outcome paradox.

First, the concept of **diminishing marginal returns** provides a key explanatory framework. Initial expansions in access and utilization effectively address unmet need and improve outcomes; however, once basic coverage is achieved, further increases yield **progressively smaller health gains** [1,5].

Second, increasing service volume may elevate the risk of **overdiagnosis and overtreatment**, particularly in systems where incentives favor activity. Excessive testing and low-value interventions can expose patients to harm without corresponding benefit, reducing overall system efficiency [3].

Third, high patient load and consultation intensity may compromise clinical encounter quality. Shorter consultations can limit assessment, weaken communication, and reduce continuity of care, thereby diminishing effectiveness [26,27,33,39].

Fourth, weak referral and gatekeeping mechanisms enable patients to bypass primary care, leading to **fragmented care, duplication, and suboptimal resource allocation** across system levels [11,13].

Taken together, these mechanisms indicate that the utilization–outcome gap is not an anomaly a pattern compatible with a health system transitioning from access expansion toward a mature, high-utilization phase. In this phase, the central challenge shifts from increasing volume to ensuring **appropriate, coordinated, and value-based care**.

2 Overuse as a Systemic Risk

High levels of healthcare utilization, when misaligned with clinical necessity, increase the risk of **systematic overuse**, posing threats to both patient safety and system efficiency. In Türkiye, persistently high consultation rates and hospital dependence create conditions in which system-level overuse and low-value care may emerge, particularly when clinical appropriateness is not routinely measured. [3,11].

Overuse manifests as unnecessary imaging, redundant testing, and avoidable procedures, driven by defensive medicine, patient expectations, and volume-based incentives. Globally, such low-value care constitutes a substantial share of healthcare inefficiency, with

significant implications for both costs and outcomes [3,11]. Beyond financial impact, overuse may expose patients to avoidable risks, including radiation exposure, procedural complications, and psychological burden associated with overdiagnosis.

Overuse should be understood not as an individual clinical failure but as a **system-level risk embedded in organizational and financial structures**. In high-intensity systems with weak gatekeeping and fragmented care, the likelihood of low-value care increases substantially [11,37,40].

The Turkish case illustrates this transition: a system that successfully expanded access now faces the challenge of **managing excess utilization and preventing low-value care**. Addressing overuse therefore requires system-level interventions, including **evidence-based practice, incentive redesign, and stronger primary care coordination** [3].

3 Structural Drivers in Türkiye

The emergence of overuse and inefficiency in Türkiye's health system reflects **interrelated structural drivers** embedded in its organizational design. These factors shape provider behavior and patient care-seeking patterns, reinforcing **high utilization and hospital-centered care**.

First, **performance-based payment systems** have driven increased service volume. While effective in improving access during early reforms, they may incentivize excessive testing, procedures, and consultations regardless of clinical necessity, thereby promoting **low-value care and overuse** [11].

Second, high patient expectations, driven by expanded access and a consumer-oriented view of healthcare, increase demand for rapid diagnostics and specialist consultations, often exceeding clinical necessity [10,32].

Third, **easy access to hospital services in the absence of strong referral and gatekeeping mechanisms** facilitates direct utilization of secondary and tertiary care. Without effective primary care coordination, patients may bypass first-contact services, leading to fragmented care pathways, duplication of services, and increased reliance on hospital-based interventions [28].

Fourth, **high physician workload** may reduce consultation time and limit comprehensive assessment. Under such constraints, clinicians may rely more on testing and referrals as time-efficient strategies, increasing the risk of **unnecessary interventions** [39,41,42].

Taken together, these structural drivers create a **reinforcing system** in which high utilization is both supply- and demand-driven. This dynamic underscores the need for **system-level reforms** that realign incentives, care pathways, and provider workload toward value-based care [37,43].

Policy Implications

1 Shift Toward Value-Based Care

A key implication is the need to transition from a **volume-driven model to value-based care**, aligning resource allocation and incentives with patient outcomes rather than service intensity.

First, implementing greater use of outcome-oriented reimbursement could help reduce reliance on activity-based incentives. Linking incentives to measurable outcomes—such as disease control, readmissions, and patient-reported outcomes—can promote effectiveness and continuity of care over service volume. Such models are increasingly adopted across OECD systems to reduce inefficiencies and low-value care [11].

Second, health systems should **reward quality over quantity**, using performance metrics that capture clinical outcomes, care coordination, patient experience, and adherence to evidence-based guidelines. This shift from throughput to quality incentivizes **appropriate, patient-centered care** and helps mitigate overuse [44].

Transitioning to **value-based care** requires not only payment reform but also supportive institutional changes, including robust data infrastructure, transparent performance measurement, and accountability mechanisms. Without these elements, value-based initiatives are unlikely to achieve their intended impact [11].

In the Turkish context, such a shift represents a critical step in addressing the utilization–outcome gap and ensuring that future improvements in health system performance are driven by **efficiency, appropriateness, and patient-centered outcomes**, rather than continued expansion of service volume.

2 Strengthening Primary Care Gatekeeping

A key policy priority is strengthening the **gatekeeping and coordinating role of primary care**, particularly within the family medicine model. Effective gatekeeping ensures care is delivered at the appropriate level, reducing unnecessary hospital use and improving system efficiency [32,35].

First, enforcing **structured referral systems** is essential to prevent bypassing of primary care. Strengthened referral pathways—supported by digital health records and appointment systems—can improve **care continuity, reduce duplication, and ensure appropriate use of specialist services** [11,28].

Second, expanding the **capacity and capability of family medicine** is essential. This includes increasing physician numbers, reducing panel sizes, and strengthening clinical problem-solving capacity. Ensuring adequate consultation time is critical, as shorter visits are associated with lower care quality and higher rates of unnecessary referrals and investigations.

In addition, strengthening primary care requires investments in **multidisciplinary team-based models**, including nurses, allied health professionals, and care coordinators, to support chronic disease management and preventive care. Such models have been shown to reduce avoidable hospitalizations and improve patient outcomes in high-performing health systems [35].

Finally, strengthening **trust in primary care** is essential, as patient preferences often favor hospital-based services. Policies should promote health literacy and system designs that position primary care as the **first and most appropriate point of contact**.

Overall, reinforcing primary care gatekeeping is a **foundational step** in addressing the utilization–outcome gap, reducing overuse, and advancing a more **efficient, coordinated, and value-based health system**.

3 Implement “Choosing Wisely”-Type Policies

A critical strategy to reduce overuse is the implementation of **“Choosing Wisely”-type policies**, which target low-value interventions through evidence-based guidance and clinician engagement [3,11].

First, the development and dissemination of **national clinical guidelines targeting low-value care** is essential. These guidelines should explicitly identify commonly overused tests, procedures, and treatments—such as unnecessary imaging, redundant laboratory testing, and inappropriate surgical interventions—and provide clear recommendations for their appropriate use. International initiatives, including the Choosing Wisely recommendations, which can support reductions in low-value care when coupled with active implementation strategies; dissemination of recommendations alone appears substantially less effective [45,46].

Second, developing and disseminating **national clinical guidelines on low-value care** is essential. These should identify commonly overused tests and interventions and provide clear recommendations for appropriate use.

The effectiveness of **Choosing Wisely-type initiatives** depends on clinician engagement and cultural change. These policies should be framed as improving **patient safety and care quality**, rather than as cost-containment. Educational interventions, audit and feedback, and peer benchmarking can support the adoption of appropriate care practices [3].

In Türkiye, such initiatives could directly address drivers of overuse and help close the **utilization–outcome gap**. Combined with payment reform and strengthened primary care, they can promote a more **efficient, evidence-based, and patient-centered health system**.

4 Improve Health Data Transparency

Improving **health data transparency** is a key enabler for reducing overuse, strengthening accountability, and advancing value-based care. Accessible data enable stakeholders to identify utilization patterns, detect inefficiencies, and monitor system performance [11].

First, **public reporting of procedure rates** can reveal unwarranted variation and potential overuse, influencing provider behavior and reducing unnecessary interventions [11,47].

Second, reporting **diagnostic utilization patterns** (e.g., imaging and laboratory use) helps identify low-value care when intensity is not linked to improved outcomes.

Third, systematic reporting of **health outcomes**, including mortality, readmissions, and patient-reported outcomes, allows assessment of whether utilization translates into meaningful health gains.

Transparency should be supported by **risk-adjusted benchmarking and comparative reporting**, ensuring fair evaluation across institutions. In Türkiye, strengthening transparency would help identify the **utilization–outcome gap** and enable targeted interventions, supporting a more **evidence-driven, accountable, and value-oriented health system**.

Conclusion

Türkiye's Health Transformation Program has substantially expanded access and utilization, addressing long-standing gaps and improving key health indicators. However, continued growth in utilization without proportional outcome gains raises concerns regarding **efficiency, appropriateness, and overuse**.

The findings suggest that Türkiye has entered a phase in which the central policy challenge is no longer access expansion alone, but optimization of the value derived from healthcare utilization. The observed divergence between high service use and more incremental gains in selected population outcomes does not establish overuse causally; rather, it highlights the need for closer assessment of clinical appropriateness, care coordination, primary care capacity, and resource allocation.

In this context, future reforms should prioritize:

- **Reducing low-value care**, particularly in high-volume service areas
- **Improving clinical appropriateness** through evidence-based guidelines and decision support
- **Aligning incentives with outcomes**, shifting from volume-driven to value-based care

More broadly, the Turkish case offers a key lesson for health systems globally: as systems mature, the focus must shift from **expanding utilization to maximizing value**. Health system success should be measured not by the quantity of care delivered, but by the **quality and value of health produced**.

This transition represents a fundamental reorientation of health policy, requiring coordinated action across **governance, financing, and service delivery**.

References

1. Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ*. 2008;27:826–842. DOI 10.1016/j.jhealeco.2007.12.002, PMID 18261812.
2. OECD. Health at a Glance 2023: OECD Indicators. Paris: OECD Publishing; 2023. DOI 10.1787/7a7afb35-en.
3. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156–168. DOI 10.1016/S0140-6736(16)32585-5, PMID 28077234,
4. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA*. 2012;307(14):1513–1516. DOI 10.1001/jama.2012.362, PMID 22419800.
5. Moreno-Serra R, Smith PC. Does progress towards universal health coverage improve population health? *Lancet*. 2012;380(9845):917–923. DOI 10.1016/S0140-6736(12)61039-3, PMID 22959388.
6. Park S, Dieleman JL, Kim R, Subramanian SV. Association of Health and Social Spending With Health Outcomes in OECD Countries. *Health Serv Res*. 2025;60(6):e14660. DOI 10.1111/1475-6773.14660, PMID 40705475,
7. Morgan DJ, Dhruva SS, Coon ER, Wright SM, Korenstein D. 2019 update on medical overuse: a review. *JAMA Intern Med*. 2019;179(11):1568–1574. doi:10.1001/jamainternmed.2019.3842.
8. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. *Ann Intern Med*. 2003;138(4):273–287. doi:10.7326/0003-4819-138-4-200302180-00006.
9. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ*. 2012;344:e3502. doi:10.1136/bmj.e3502.
10. Saini V, Brownlee S, Elshaug AG, Glasziou P, Heath I. Addressing overuse and underuse around the world. *Lancet*. 2017;390(10090):105–107. DOI 10.1016/S0140-6736(16)32573-9, PMID 28077230.8
11. OECD. Tackling Wasteful Spending on Health. Paris: OECD Publishing; 2017. DOI 10.1787/9789264266414-en.9
12. Studdert DM, Mello MM, Sage WM, et al. Defensive medicine among high-risk specialist physicians. *JAMA*. 2005;293(21):2609–2617. DOI 10.1001/jama.293.21.2609, PMID 15928282.10
13. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q*. 2005;83(3):457–502. DOI 10.1111/j.1468-0009.2005.00409.x, PMID 16202000.11
14. Papanicolas I, Smith PC, editors. Health System Performance Comparison: An Agenda for Policy, Information and Research. Open University Press; 2013. 360 p. ISBN 9780335247264 / 9780335247271.12
15. World Bank. Life expectancy at birth, total (years) [Internet]. World Development Indicators. Washington (DC): World Bank; [cited 2026 Sep 18]. Available from: World Bank – Life expectancy at birth, total (years)
16. World Bank. Mortality rate, infant (per 1,000 live births) [Internet]. World Development Indicators. Washington (DC): World Bank; [cited 2026 Sep 18]. Available from: World Bank – Infant mortality rate
17. Health Statistics Yearbook. Ankara: T.C. Sağlık Bakanlığı / Sağlık Bilgi Sistemleri Genel Müdürlüğü. 15
18. Atun R, Aydın S, Chakraborty S, Sümer S, Aran M, Gürol I, et al. Universal health coverage in Turkey: Enhancement of equity. *Lancet*. 2013;382(9886):65–99. DOI 10.1016/S0140-6736(13)61051-X, PMID 23810020.16
19. Turkish Statistical Institute (TÜİK). Life Satisfaction Survey, 2023 [Internet]. Ankara: TÜİK; 2024 Feb 16 [cited 2026 Sep 20]. Available from: <https://veriportali.tuik.gov.tr/tr/press/49692>
20. Turkish Statistical Institute (TÜİK). Life Satisfaction Survey, 2015 [Internet]. Ankara: TÜİK; 2016 Feb 23 [cited 2026 Sep 20]. Available from: <https://veriportali.tuik.gov.tr/tr/press/21518>
21. Organisation for Economic Co-operation and Development (OECD). Rethinking Health System Performance Assessment: A Renewed Framework. OECD Health Policy Studies. Paris: OECD Publishing; 2024. doi:10.1787/107182c8-en.
22. Y. Papanicolas I, Rajan D, Karanikolos M, Soucat A, Figueras J, editors. Health system performance assessment: a framework for policy analysis. Geneva: World Health Organization; 2022. Health Policy Series No. 57.
23. Republic of Türkiye Ministry of Health, General Directorate of Health Information Systems. Health Statistics Yearbook 2023 [Internet]. Ankara: Ministry of Health; 2025 [cited 2026 Sep 20]. Available from: <https://sbsgm.saglik.gov.tr/TR-107089/health-statistics-yearbook-2023.html>
24. Federal Reserve Bank of St. Louis. Infant mortality rate for Turkey [Internet]. FRED, Federal Reserve Bank of St. Louis; [cited 2026 Sep 18]. Available from: <https://fred.stlouisfed.org/series/SPDYNIMRTINTUR>
25. OECD. Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance? OECD Health Policy Studies. Paris: OECD Publishing; 2014. doi:10.1787/9789264216594-en.
26. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*. 2017;7:e017902. DOI 10.1136/bmjopen-2017-017902, PMID 29118053.18
27. Zyzanski SJ, Stange KC, Langa D, Flocke SA. Trade-offs in high-volume primary care practice. *J Fam Pract*. 1998;46(5):397–402.
28. European Observatory on Health Systems and Policies and World Health Organization. Regional Office for Europe. (2022). Health systems in action: Türkiye: 2022 edition. World Health Organization. Regional Office for Europe19
29. Federal Reserve Bank of St. Louis. Life expectancy at birth, total for Turkey [Internet]. FRED; 2026 [cited 2026 Sep 20]. Available from: <https://fred.stlouisfed.org/data/SPDYNLE00INTUR>

30. Hensher M, Blizzard L, Campbell J, Canny B, Zimitat C, Palmer A. Diminishing marginal returns and sufficiency in health-care resource use: an exploratory analysis of outcomes, expenditure, and emissions. *Lancet Planet Health*. 2024;8(10):e744-e753. doi:10.1016/S2542-5196(24)00207-9.
31. Joumard I, André C, Nicq C, Chatal O. Health status determinants: lifestyle, environment, health care resources and efficiency. OECD Economics Department Working Papers. 2008;(627). Paris: OECD Publishing. doi:10.1787/240858500130.
32. Organisation for Economic Co-operation and Development (OECD). Realising the Potential of Primary Health Care. OECD Health Policy Studies. Paris: OECD Publishing; 2020. doi:10.1787/a92adee4-en.
33. Wilson A, Childs S. The relationship between consultation length, process and outcomes in general practice: a systematic review. *Br J Gen Pract*. 2002;52(485):1012-1020.
34. Tatar M, Mollahaliloğlu S, Şahin B, Aydın S, Maresso A, Hernández-Quevedo C. Turkey: health system review. *Health Syst Transit*. 2011;13(6):1-186.
35. Rosano A, Loha CA, Falvo R, van der Zee J, Ricciardi W, Guasticchi G, et al. The relationship between avoidable hospitalization and accessibility to primary care: a systematic review. *Eur J Public Health*. 2013;23:356–360. DOI 10.1093/eurpub/cks053, PMID 22645236.20
36. Kringos DS, Boerma WG, Hutchinson A, van der Zee J, Groenewegen PP. The breadth of primary care: a systematic literature review of its core dimensions. *BMC Health Serv Res*. 2010;10:65. DOI 10.1186/1472-6963-10-65, PMID 20226084. 21
37. OECD/The Health Foundation. How Do Health System Features Influence Health System Performance? Paris: OECD Publishing; 2025. doi:10.1787/7b877762-en.
38. Walt, Gilson. *Health Policy Plan*. 1994;9:353–370. DOI 10.1093/heapol/9.4.353, PMID 10139469..22
39. León-García M, Wieringa TH, Espinoza Suárez NR, et al. Does the duration of ambulatory consultations affect the quality of healthcare? A systematic review. *BMJ Open Qual*. 2023;12(4):e002311. doi:10.1136/bmjopen-2023-002311.
40. Ellen ME, Wilson MG, Vélez M, Shach R, Lavis JN, Grimshaw JM, et al. Addressing overuse of health services in health systems: a critical interpretive synthesis. *Health Res Policy Syst*. 2018;16(1):48. doi:10.1186/s12961-018-0325-x.
41. Sirovich BE, Woloshin S, Schwartz LM. Too little? Too much? Primary care physicians' views on US health care: a brief report. *Arch Intern Med*. 2011;171(17):1582-1585. doi:10.1001/archinternmed.2011.437.
42. Nguyen MT, Honcharov V, Ballard D, Satterwhite S, McDermott AM, Sarkar U. Primary Care Physicians' Experiences With and Adaptations to Time Constraints. *JAMA Netw Open*. 2024 Apr 1;7(4):e248827. doi: 10.1001/jamanetworkopen.2024.8827.
43. Hämmerli M, Milstein R, Mueller M. Incentivising patient pathways in outpatient care: a review of gatekeeping and cost-sharing policies across the OECD. *OECD Health Working Papers*. 2026;(196). Paris: OECD Publishing; 2026. doi:10.1787/349dcf1c-en.
44. Lindner L, Lorenzoni L. Innovative providers' payment models for promoting value-based health systems: start small, prove value, and scale up. *OECD Health Working Papers*. 2023;(154). Paris: OECD Publishing; 2023. doi:10.1787/627fe490-en.
45. Cliff BQ, Avanceña ALV, Hirth RA, Lee SYD. The impact of Choosing Wisely interventions on low-value medical services: a systematic review. *Milbank Q*. 2021;99(4):1024-1058. doi:10.1111/1468-0009.12531.
46. Grimshaw JM, Patey AM, Kirkham KR, Hall A, Dowling SK, Rodondi N, et al. De-implementing wisely: developing the evidence base to reduce low-value care. *BMJ Qual Saf*. 2020;29(5):409-417. doi:10.1136/bmjqs-2019-010060.
47. Prang KH, Maritz R, Sabanovic H, Dunt D, Kelaher M. Mechanisms and impact of public reporting on physicians and hospitals' performance: a systematic review (2000–2020). *PLoS One*. 2021;16(2):e0247297. doi:10.1371/journal.pone.0247297.

Data rich, meaning poor

Hakan Yaman

Author bio:

Dr. Hakan Yaman is a family physician and former academic based in Türkiye. His work focuses on primary care systems, health policy, and medical humanities.

Modern healthcare prides itself on precision, standardisation, and efficiency. Yet, paradoxically, both patients and clinicians increasingly experience care as fragmented, impersonal, and exhausting. Nowhere is this tension more visible than in primary care, where continuity of care has become more difficult to sustain in increasingly fragmented systems [1].

In Türkiye, family medicine has undergone rapid and large-scale reform over the past two decades. Access has improved, preventive services have expanded, and key public health indicators have advanced (2). But alongside these successes, a quieter problem has emerged: a growing gap between what is measured and what actually matters to patients. This is not merely a technical issue, it is a problem of meaning (3).

Primary care systems today are increasingly structured around performance indicators, clinical pathways, and digital prompts. These tools are valuable, but they subtly reshape the clinician's role—from someone who interprets a patient's experience to someone who processes clinical data (4).

Having more data about a patient is not the same as understanding their situation.

In everyday practice in Türkiye, this shift is tangible. High patient volumes, short consultations, and checklist-driven encounters mean that clinicians often meet formal requirements of care while missing the patient's lived reality. The result is a paradox: data-rich, but meaning-poor medicine.

Primary care has traditionally been the domain of generalism—the ability to integrate biomedical knowledge with the patient's life context. However, current systems prioritise what can be standardised and measured, while undervaluing what must be interpreted.

A hermeneutic approach offers a necessary corrective. It shifts the focus from “What is the diagnosis?” to “What does this illness mean for this person?” It asks clinicians to engage with narratives, fears, and expectations, not as optional extras, but as central components of care (5).

Consider a common scenario: a patient presents repeatedly with non-specific pain. Investigations are normal, and guidelines suggest reassurance or referral. Yet a more interpretive approach may reveal caregiving burden, financial stress, or unresolved trauma. Without engaging these dimensions, care becomes repetitive, inefficient, and ultimately ineffective.

There is a persistent assumption that meaning-making takes time and reduces efficiency. In reality, the opposite may be true. A slightly longer consultation that uncovers meaning may reduce unnecessary investigations, referrals, and repeat visits. Meaning is not a luxury, it is a form of system efficiency (5).

This has important implications for Türkiye. High utilisation in primary care is often interpreted as a problem of access or patient behaviour. But it may also reflect a failure to address the meaning behind symptoms (5). When care does not resonate with patients' lived realities, they return—not because the system is accessible, but because it is insufficient.

The rapid digitalisation of healthcare further intensifies this tension. Algorithmic decision-making and remote consultations risk weakening the relational space between patient and clinician. Yet it is precisely within this space that care becomes meaningful.

If primary care is reduced to a series of transactions, it may remain efficient, but it will cease to be care.

Türkiye's family medicine system has achieved significant structural success (2). The next phase of reform should not only refine metrics or expand services, but restore the interpretive dimension of care.

This does not require abandoning evidence-based medicine. It requires integrating evidence with interpretation, ensuring that clinical decisions are not only scientifically valid, but meaningful to the person in front of us.

Because ultimately, healthcare is not only about solving problems, it is about making sense of them.

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

References:

1. Freeman GK, Olesen F, Hjortdahl P. Continuity of care: an essential element of modern general practice? *Fam Pract.* 2003 Dec;20(6):623-7. doi: 10.1093/fampra/cmz601. PMID: 14701883.
2. Yaman H, Güneş ED. Family practice in Turkey: Observations from a pilot implementation. *Scand J Prim Health Care.* 2016;34(1):81-2. doi: 10.3109/02813432.2016.1144432. Epub 2016 Feb 19. PMID: 26893201; PMCID: PMC4911021.
3. Coulter A. Measuring what matters to patients. *BMJ.* 2017 Feb 20;356:j816. doi: 10.1136/bmj.j816. PMID: 28219884.
4. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *BMJ.* 2016 Aug 4;354:i4060. doi: 10.1136/bmj.i4060. PMID: 27492602; PMCID: PMC4975019.
5. Shah R, Clarke R, editors. *Finding Meaning in Healthcare: Looking Through the Hermeneutic Window.* Abingdon: Routledge; 2026.

Author bio:

Dr. Hakan Yaman is a family physician and former academic based in Türkiye. His work focuses on primary care systems, health policy, and medical humanities.



International Institute on Ageing
United Nations - Malta

MEAMA
*Middle-East Academy
for Medicine of Ageing*



Postgraduate Course in Dementia Care

Middle East Academy for Medicine of Ageing (MEAMA)
In Collaboration with MENAAA & INIA - United Nations Malta

Executive Introduction

The Postgraduate Course in Dementia Care is a structured postgraduate program developed by MEAMA to address the growing burden of dementia across the Middle East and North Africa (MENA). The program integrates evidence-based medicine, interdisciplinary collaboration, and culturally sensitive care models to strengthen regional capacity in dementia diagnosis, management, and leadership.

Program Vision

The program aims to:

- Promote person-centered, ethical, and evidence-based dementia care
- Strengthen regional capacity in early detection, diagnosis, and management
- Improve care pathways across outpatient, hospital, community, and long-term care settings
- Empower interdisciplinary teams to lead change in dementia services

Curriculum Structure

The Course includes 10 modules delivered across four sessions (two virtual and two hybrid). Core modules cover epidemiology, diagnosis, medical and non-pharmacological management, ethics, palliative care, care pathways, and caregiver support.

Target Audience

The Diploma is designed for:

- Physicians (primary care, geriatrics, neurology, psychiatry, internal medicine)
- Nurses and allied health professionals
- Social workers and psychologists
- Occupational and physical therapists
- Healthcare administrators
- Academic faculty and trainees

It supports both clinical advancement and leadership development in dementia care.

Certification

Participants who successfully complete all modules receive the Postgraduate Certificate in Dementia Care issued under the auspices of MEAMA in collaboration with MENAAA and INIA.

Certificate in Dementia Care

Middle East Academy for Medicine of Ageing

> Comprehensive Training for Dementia Professionals <

In Collaboration with



Program Structure

4 Sessions - 10 Core Modules



Core Modules



Earn Your Postgraduate Certificate



Cognitive Testing: Bedside Screening Tools and Formal Neuropsychological Assessment - A Comprehensive Academic Review

A Abyad

Correspondence:

A. Abyad, MD, MPH, MBA, DBA, AGSF

Consultant, Internal Medicine and Geriatrics, Dar Al Shifa Hospital, Kuwait

Chairman, Middle-East Academy for Medicine of Aging.

President, Middle East & North Africa Association on Aging & Alzheimer's

Coordinator, Middle-East Primary Care Research Network

Coordinator, Middle-East Network on Aging

Email: aabyad@cyberia.net.lb

Received: August 2026. Accepted: September 2026; Published: October 2026.

Citation: A. Abyad. Cognitive Testing: Bedside Screening Tools and Formal Neuropsychological Assessment

A Comprehensive Academic Review. World Family Medicine. October 2026; 24(7): 36 - 54 DOI: 10.5742/MEWFM.2026.241794

Abstract

Cognitive testing constitutes an indispensable component of contemporary clinical neuroscience, psychiatry, geriatrics, and primary care, providing the empirical foundation upon which diagnoses of mild and major neurocognitive disorder are established, staged, and monitored. This review provides a comprehensive, evidence-based examination of the two principal tiers of cognitive assessment employed in clinical practice: brief bedside cognitive screening instruments, exemplified by the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA), and formal neuropsychological testing conducted by licensed neuropsychologists. Drawing on diagnostic accuracy meta-analyses, clinical practice guidelines, and recent advances in digital and biomarker-based assessment, the review synthesizes evidence published predominantly between 2020 and 2026 alongside foundational diagnostic frameworks. Diagnostic criteria tables, comparative accuracy tables, and a stepwise clinical algorithm are presented to orient clinicians toward evidence-based test selection. The review further addresses performance validity assessment, differential diagnosis across dementia subtypes, cultural and educational bias in screening instruments, and emerging digital and blood-based biomarker technologies that

are reshaping the diagnostic pathway. The review concludes that bedside screening and formal neuropsychological testing occupy complementary, non-substitutable positions within a stepped, biologically informed model of cognitive evaluation, and that clinicians must select instruments deliberately according to clinical context, patient characteristics, and diagnostic yield.

Keywords: cognitive testing, screening tools, neuropsychological assessment

Introduction

Cognitive impairment represents one of the most consequential and rapidly growing challenges confronting global health systems. The Alzheimer's Association (2024) estimates that millions of individuals in the United States alone are living with Alzheimer's disease and related dementias, a figure projected to rise substantially as the population ages. Rajan et al. (2021) similarly project that the number of individuals with clinical Alzheimer's disease and mild cognitive impairment (MCI) in the United States will increase considerably between 2020 and 2060. On a global scale, Song et al. (2023) report that the pooled prevalence of MCI among older adults exceeds fifteen percent, with substantial regional variation attributable to demographic, educational, and methodological factors. These epidemiological trends underscore the urgency of accurate, accessible, and psychometrically sound cognitive assessment across the full continuum of clinical settings, from primary care offices to specialized memory clinics.

Cognitive functioning is not a unitary construct; rather, it comprises multiple interdependent domains, including episodic memory, working memory, executive function, language, visuospatial processing, attention, and processing speed. Because impairment can be circumscribed to a single domain or distributed across several, and because the pattern of impairment often carries substantial diagnostic significance, no single test can adequately characterize the entirety of an individual's cognitive profile. Clinical practice has therefore evolved a tiered, stepped approach to cognitive assessment. At the first tier, brief bedside screening instruments, administered in a matter of minutes by physicians, nurses, or other trained clinical personnel, serve a triage function: they flag the possible presence of cognitive impairment, provide a coarse index of severity, and inform decisions about further referral. At the second tier, formal neuropsychological testing, conducted by licensed neuropsychologists over several hours, provides a granular, domain-specific profile of cognitive strengths and weaknesses, integrates behavioural observation and collateral history, and supports differential diagnosis, prognosis, and treatment planning with a level of precision that screening instruments cannot achieve.

The clinical stakes of accurate cognitive assessment have grown substantially in recent years with the regulatory approval of disease-modifying therapies for early Alzheimer's disease, including lecanemab (van Dyck et al., 2023) and donanemab (Sims et al., 2023), both of which require biomarker confirmation of amyloid pathology and careful staging of cognitive-functional status prior to treatment initiation. This therapeutic context has elevated the diagnostic precision demanded of the cognitive assessment pathway, reinforcing the necessity of a well-coordinated stepped model that begins with screening and, where indicated, proceeds to formal neuropsychological evaluation and biomarker-based

diagnostic workup (Jack et al., 2024). This paper reviews the evidence base underlying bedside cognitive screening tools and formal neuropsychological testing, situates both within the broader diagnostic and nosological framework of neurocognitive disorders, and proposes an evidence-based clinical algorithm to guide assessment across the continuum of care.

Nosological and Diagnostic Framework of Cognitive Impairment

Before examining specific assessment instruments, it is necessary to situate cognitive testing within the broader diagnostic frameworks that determine how test results are interpreted and applied clinically. Two frameworks predominate in contemporary practice: the syndromic, function-based criteria articulated in the fifth edition, text revision, of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR; American Psychiatric Association, 2022), and the biologically anchored framework developed by the National Institute on Aging and Alzheimer's Association (NIA-AA), most recently revised by Jack et al. (2024).

1 DSM-5-TR Neurocognitive Disorders

The DSM-5-TR reorganizes the older construct of dementia under the broader heading of neurocognitive disorder (NCD), which is subdivided into mild NCD and major NCD (American Psychiatric Association, 2022). This nomenclature, first introduced in DSM-5 and retained in the text revision, reflects an effort to capture the full continuum of cognitive decline, including its earliest clinically meaningful stages, rather than restricting diagnosis to the more severe, function-impairing presentation historically labeled dementia. Diagnosis under this framework requires evidence of cognitive decline from a previous level of performance in one or more of six specified domains, complex attention, executive function, learning and memory, language, perceptual-motor function, and social cognition, established through standardized neuropsychological testing wherever possible, or, when formal testing is unavailable, through clinical assessment. Critically, the DSM-5-TR criteria distinguish mild from major NCD according to the degree of functional impact: major NCD requires that cognitive deficits interfere with independence in everyday activities, necessitating compensation or assistance, whereas mild NCD permits preserved independence, albeit with greater effort, compensatory strategies, or accommodation (American Psychiatric Association, 2022).

Etiological subtyping is appended to the core NCD diagnosis, with specified codes for NCD due to Alzheimer's disease, frontotemporal degeneration, Lewy body disease, vascular disease, traumatic brain injury, substance or medication use, HIV infection, prion disease, Parkinson's disease, Huntington's disease, another medical condition, multiple etiologies, or unknown etiology (American Psychiatric Association, 2022). This etiological specification underscores the centrality of differential

diagnosis to the neurocognitive assessment process and highlights why formal neuropsychological testing, capable of generating domain-specific profiles associated with particular etiologies, occupies an essential role beyond the detection function served by bedside screening.

2 The NIA-AA Biological Framework and 2024 Revised Criteria

In parallel with the DSM's syndromic approach, the NIA-AA has developed a biologically anchored diagnostic framework for Alzheimer's disease that has evolved considerably over the past decade and a half. The 2011 NIA-AA workgroups articulated separate criteria for the preclinical, MCI, and dementia stages of Alzheimer's disease (Albert et al., 2011; McKhann et al., 2011), establishing Alzheimer's disease as a clinical-biomarker construct. Building on the amyloid-tau-neurodegeneration (ATN) framework introduced in 2016, the 2018 NIA-AA research framework operationalized biomarkers for staging Alzheimer's disease across its full continuum, from asymptomatic biomarker positivity through overt dementia.

Most recently, Jack et al. (2024) published revised criteria for the diagnosis and staging of Alzheimer's disease on behalf of an Alzheimer's Association workgroup, incorporating blood-based biomarkers into the classification system for the first time and introducing a two-dimensional biological-clinical staging matrix. A central principle of the revised criteria is that Alzheimer's disease should be defined biologically rather than by clinical syndrome alone, consistent with practice in other areas of medicine such as oncology (Jack et al., 2024). Under this framework, the disease is conceptualized as a continuum, beginning with amyloid deposition in asymptomatic individuals (biological stage A, clinical stage 1), progressing through tau deposition and subtle cognitive impairment (biological stage B, clinical stage 2), and culminating in progressive neurodegeneration and dementia. Application of these criteria in drug development and clinical practice contexts has been discussed extensively, with commentators noting that although the criteria offer objective, biologically grounded measures for diagnosis and staging, their implementation depends heavily on the availability and accuracy of biomarkers, particularly increasingly accessible plasma-based assays (Jack et al., 2024/2025). The 2024 criteria have generated active discussion regarding their applicability to non-Alzheimer's dementia syndromes, including dementia with Lewy bodies, for which the relationship between amyloid co-pathology and clinical staging remains an area of ongoing investigation (Abdelnour, 2025).

The coexistence of the syndromic DSM-5-TR framework and the biological NIA-AA framework has direct implications for cognitive testing. Bedside screening and formal neuropsychological testing remain the primary instruments through which the syndromic, functional dimension of diagnosis is established, while biomarker studies (amyloid positron emission tomography, cerebrospinal fluid analysis, and increasingly, blood-based

biomarkers) establish the biological dimension. The two are complementary: a biologically positive amyloid scan in a cognitively normal individual does not, by itself, constitute a diagnosis of Alzheimer's dementia in the clinical sense, and a diagnosis of major NCD requires demonstration of functional impairment that biomarkers alone cannot establish (Shaughnessy & Weintraub, 2024).

Table 1. DSM-5-TR Diagnostic Criteria: Mild versus Major Neurocognitive Disorder

Criterion	Mild Neurocognitive Disorder	Major Neurocognitive Disorder
Cognitive decline	Modest decline from previous level in one or more cognitive domains, based on concern of the individual, informant, or clinician, and objective evidence of impairment (typically test performance in the 1–2 SD below appropriate norms range)	Significant decline from previous level in one or more cognitive domains, with objective impairment typically 2 or more SD below appropriate norms
Functional impact	Cognitive deficits do not interfere with capacity for independence in everyday activities (e.g., instrumental activities of daily living remain intact, though greater effort or compensatory strategies may be required)	Cognitive deficits interfere with independence in everyday activities, necessitating assistance
Delirium exclusion	Deficits do not occur exclusively in the context of delirium	Deficits do not occur exclusively in the context of delirium
Other mental disorder exclusion	Deficits not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia)	Deficits not better explained by another mental disorder
Etiological specifier	Specified where possible (e.g., due to Alzheimer's disease, due to Lewy body disease, due to vascular disease, due to unknown etiology)	Specified where possible, with additional severity specifier (mild, moderate, severe) based on functional dependence
Typical assessment basis	Standardized neuropsychological testing preferred; clinical assessment when testing unavailable	Standardized neuropsychological testing preferred; clinical assessment when testing unavailable

Note. Adapted from diagnostic criteria described in American Psychiatric Association (2022).

Bedside Cognitive Screening Tools

Bedside cognitive screening instruments occupy the first tier of the stepped cognitive assessment pathway. They are characterized by brevity (typically five to fifteen minutes of administration time), minimal training requirements, and applicability across a wide range of clinical settings, including primary care offices, emergency departments, hospital wards, and community health screenings. The American Academy of Neurology practice guideline update specifies that when screening or assessing for MCI, clinicians should employ validated assessment tools, noting that a range of instruments demonstrate acceptable diagnostic accuracy without clear superiority of any single instrument (Petersen et al., 2018). Critically, the guideline emphasizes that because brief cognitive screening instruments are typically calibrated to maximize sensitivity at some cost to specificity, patients who screen positive should undergo further evaluation, potentially including formal neuropsychological testing with normative interpretation, before a diagnosis is confirmed (Petersen et al., 2018). This principle, that screening identifies candidates for further evaluation rather than establishing a definitive diagnosis, constitutes the conceptual foundation of the entire stepped assessment model and is a recurring theme throughout the evidence reviewed in this section.

1 The Mini-Mental State Examination (MMSE)

The MMSE, originally developed by Folstein, Folstein, and McHugh (1975), remains among the most extensively studied and internationally disseminated cognitive screening instruments in clinical use. The instrument comprises eleven items assessing orientation to time and place, registration of three words, attention and calculation (commonly operationalized through serial subtraction), delayed recall of the three registered words, language (naming, repetition, comprehension of a three-stage command, reading, writing), and visuoconstructional praxis (intersecting pentagon copy), yielding a total score ranging from zero to thirty points (Folstein et al., 1975).

The diagnostic accuracy of the MMSE has been the subject of extensive meta-analytic scrutiny. A Cochrane systematic review examining the MMSE for detection of dementia in community and primary care populations aged sixty-five and older found that, at a cutoff of twenty-four, the instrument achieved pooled sensitivity of approximately 0.85 and specificity of approximately 0.90 in community settings; at a cutoff of twenty-five, pooled sensitivity rose to approximately 0.87 with reduced specificity of approximately 0.82 (Creavin et al., 2016). Notably, when accuracy estimates were adjusted for level of education, sensitivity increased to approximately 0.97 while specificity fell to approximately 0.70, illustrating the substantial confounding influence of educational attainment on MMSE performance (Creavin et al., 2016). A separate Cochrane review examining the MMSE's utility in predicting conversion from MCI to dementia found highly variable accuracy, with sensitivities for conversion to Alzheimer's disease dementia ranging from twenty-seven to eighty-nine percent and specificities ranging from thirty-

two to ninety percent across included studies, leading the reviewers to conclude that there is insufficient evidence to support a substantial role for the MMSE as a stand-alone, single-administration predictor of conversion from MCI to dementia (Arevalo-Rodriguez et al., 2021).

Recent work has further probed the limitations of relying on the MMSE in isolation. Kim et al. (2025), employing a deep learning classifier trained on clinical data, demonstrated that a model relying solely on MMSE scores achieved an area under the receiver operating characteristic curve of 0.872 for predicting Clinical Dementia Rating category, whereas a composite model incorporating additional cognitive assessments achieved a significantly higher area under the curve of 0.934, indicating meaningfully improved diagnostic precision when the MMSE is supplemented with additional cognitive measures (Kim et al., 2025). The same investigation raised concerns regarding dementia overdiagnosis among younger, more highly educated individuals when the MMSE is used as the sole diagnostic instrument, reinforcing longstanding concerns regarding the interaction between educational attainment and MMSE performance (Kim et al., 2025).

Beyond its established limitations with respect to education and cultural background, the MMSE has been repeatedly criticized for relatively poor sensitivity to executive dysfunction and to the subtle deficits characteristic of MCI, for pronounced ceiling effects among highly educated respondents, and for the historical restriction of its use owing to copyright licensing requirements, a factor that contributed substantially to the increasing clinical adoption of alternative, freely available instruments such as the MoCA (Nasreddine et al., 2005).

2 The Montreal Cognitive Assessment (MoCA)

The MoCA was developed by Nasreddine et al. (2005) specifically to address the MMSE's reduced sensitivity to MCI. The instrument evaluates eight cognitive domains, visuospatial and executive function (via a trail-making task, a cube copy, and a clock-drawing task), naming, memory (five-word learning across two trials with delayed recall), attention (digit span, target detection, serial subtraction), language (sentence repetition and phonemic fluency), abstraction, delayed recall, and orientation, within a total administration time of approximately ten to fifteen minutes, yielding a score out of thirty points, with one additional point awarded to individuals with twelve years of education or fewer to partially correct for educational bias (Nasreddine et al., 2005).

The originally proposed optimal cutoff of twenty-six has been substantially revised by subsequent meta-analytic evidence. Islam et al. (2023), in a systematic review and meta-analysis pooling thirteen studies encompassing 2,158 participants, found that a cutoff of less than twenty-three maximized the combined sensitivity and specificity for detecting MCI, yielding pooled sensitivity of 73.5% and specificity of 91.3%; at the traditionally recommended cutoff of less than twenty-six, sensitivity rose to 83.8% but specificity fell to 70.8%, reflecting a substantial trade-off

between case detection and false-positive classification (Islam et al., 2023). These findings are broadly consistent with earlier systematic reviews and diverge notably from the original 2005 validation study, a discrepancy the authors attribute in part to differences in sample size and population characteristics between the original validation cohort and subsequent independent replications (Islam et al., 2023).

The MoCA's cross-cultural and cross-linguistic performance has received considerable recent attention. Systematic reviews addressing the instrument's validity in culturally diverse populations, including Spanish-speaking populations of Latin America and the Caribbean, have documented substantial heterogeneity in reported cutoffs, translation and adaptation practices, and psychometric reporting quality, with educational level consistently identified as a source of bias across populations studied (Cornea, Simu, & Rosca, 2022). The MoCA has also been extended well beyond its original purpose of detecting MCI in older adults to encompass cognitive screening in numerous other neurological and psychiatric populations, including subarachnoid hemorrhage, where systematic review evidence indicates good internal consistency and sensitivity to cognitive impairment in the subacute and chronic post-hemorrhage phases (Cornea, Simu, & Rosca, 2022).

Comparative studies incorporating remote and digital assessment modalities have further situated the MoCA within the evolving diagnostic landscape. Thompson et al. (2023) compared remote smartphone-based cognitive tasks, in-clinic digital tablet-based assessments, and the MoCA in their ability to distinguish cerebral amyloid status among cognitively healthy older adults, finding that certain digital episodic memory tasks achieved discriminative accuracy modestly exceeding that of the MoCA (area under the curve of 0.77 versus 0.74, respectively), suggesting that digital cognitive assessment may offer incremental sensitivity to very early, preclinical stages of Alzheimer's pathology beyond what conventional bedside screening captures (Thompson et al., 2023).

3 The Mini-Cog

The Mini-Cog, developed by Borson, Scanlan, Chen, and Ganguli (2003), combines a three-item delayed word recall task with a clock-drawing test, yielding administration times of approximately three to five minutes, making it among the briefest validated cognitive screening instruments in widespread use. In its original validation study, conducted within a population-based epidemiological sample, the Mini-Cog demonstrated sensitivity and specificity for dementia detection (76% and 89%, respectively) comparable to the MMSE at a cutpoint of twenty-five (79% and 88%, respectively) and to a conventional neuropsychological battery (75% and 90%, respectively) (Borson et al., 2003). However, a subsequent Cochrane systematic review evaluating the Mini-Cog for dementia detection in primary care settings identified substantial heterogeneity across studies and methodological limitations in all but one included investigation; the single low-risk-of-bias

study reported sensitivity of 76% and specificity of 73%, leading the Cochrane reviewers to conclude that there is presently insufficient evidence to routinely recommend the Mini-Cog as a standalone dementia screening test in primary care, notwithstanding its practical appeal (Seitz et al., 2021). The instrument's brevity nonetheless renders it a pragmatic option in time-constrained or resource-limited settings, provided that positive screens are followed by further evaluation.

4 The Saint Louis University Mental Status (SLUMS) Examination

The SLUMS examination, developed within the Veterans Affairs system, is an eleven-item, thirty-point instrument assessing orientation, memory, attention, and executive function through tasks including animal naming, digit span, figure recognition, clock drawing, and size differentiation, with administration requiring approximately seven minutes (Spencer, Noyes, Bair, & Ransom, 2022). The instrument was developed in part to address perceived limitations of the MMSE, including its heavy reliance on orientation items and its comparatively limited sensitivity among more highly educated examinees. A systematic review of the SLUMS's psychometric properties concluded that the instrument demonstrates adequate validity, sensitivity, and specificity for detecting cognitive impairment, while noting the continued need for more diverse normative samples and further test-retest reliability data before its use can be recommended without qualification (Spencer et al., 2022). Reported cutoffs, adjusted for educational attainment, generally classify scores of twenty-seven to thirty as within normal limits, twenty-one to twenty-six as suggestive of mild neurocognitive disorder, and scores at or below twenty as suggestive of major neurocognitive disorder, though sensitivity and specificity at these cutoffs vary appreciably across studies and populations (Spencer et al., 2022). A notable practical advantage of the SLUMS relative to the MMSE and MoCA is that it resides in the public domain and may be administered without licensing fees or mandatory training modules, a consideration of increasing relevance as healthcare systems seek to expand access to cognitive screening. Recent comparative work has begun to validate digital adaptations against the SLUMS in real-world clinical samples, reporting concurrent validity supportive of digital platforms as a scalable complement to traditional paper-based administration (Huynh, Patterson, & Huang, 2025).

5 Culture-Fair and Education-Independent Screening Instruments

A substantial and growing body of literature has documented that widely used cognitive screening instruments, including the MMSE and MoCA, are subject to systematic bias related to educational attainment, literacy, language proficiency, and cultural background. A comprehensive systematic review examining culture-fair cognitive screening tools identified eleven distinct instruments developed specifically to minimize such bias, including the Rowland Universal Dementia Assessment Scale (RUDAS), the Community Screening Instrument for Dementia (CSI-D), and the Visual Cognitive Assessment

Test (VCAT) (Chithiramohan et al., 2024). The review found that the RUDAS achieved diagnostic accuracy comparable to the MMSE (area under the curve ranging from 0.62 to 0.93) while demonstrating superior specificity, and that the RUDAS, CSI-D, and VCAT were minimally influenced by educational level, with the VCAT proving particularly useful for non-verbal patients or those with limited education or language proficiency (Chithiramohan et al., 2024). These findings carry direct implications for clinical practice in linguistically and culturally diverse populations, where reliance on standard MMSE or MoCA cutoffs without adaptation risks both under- and over-diagnosis of cognitive impairment.

6 Delirium Screening: The Confusion Assessment Method and Related Instruments

Although not a general cognitive screening tool in the conventional sense, delirium detection instruments occupy an essential adjunctive role within the bedside cognitive assessment armamentarium, as unrecognized delirium can profoundly confound the interpretation of cognitive screening results obtained in acute care settings. The Confusion Assessment Method (CAM), developed by Inouye and colleagues (1990), remains the most widely used and internationally validated delirium screening algorithm, evaluating four cardinal features: acute onset and fluctuating course, inattention, disorganized thinking, and altered level of consciousness. Diagnosis of delirium by CAM requires the presence of the first two features plus either of the remaining two (Inouye et al., 1990). Comparative studies of ultrabrief delirium screening instruments, including the 4 'A's Test (4AT) and modified CAM protocols adapted for emergency department use, have generally found favorable diagnostic accuracy and rapid administration times suitable for use by non-specialist personnel (Muser et al., 2022). Large-scale implementation research indicates that structured protocols can substantially increase delirium screening rates in hospital settings; one quality improvement study documented an increase in the proportion of hospital admissions screened for delirium from 74% to 98.7% over a four-year period following protocol implementation, although adherence to recommended daily screening frequency remained limited even after implementation, illustrating persistent challenges in translating validated instruments into consistent bedside practice (Penfold et al., 2024).

7 Digital and Remote Cognitive Screening

The proliferation of digital health technologies has substantially expanded the modalities through which bedside cognitive screening can be conducted, with particular acceleration following the COVID-19 pandemic's disruption of in-person clinical services. Owens et al. (2020) documented the rapid implementation of remote memory clinic services during and after the pandemic, demonstrating the feasibility of adapting cognitive assessment pathways to telehealth delivery. Subsequent validation research has examined the reliability of self-administered, remotely delivered digital cognitive assessments relative to clinician-administered or in-person

equivalents. Kochan et al. (2022) evaluated the reliability, validity, and user experience of remote unsupervised computerized neuropsychological assessments among community-living older adults, providing empirical support for the feasibility of unsupervised digital testing in this population. Morrissey, Gillings, and Hornberger (2024) similarly reported the feasibility and reliability of online relative to in-person cognitive testing among healthy older adults. Comparative work examining agreement between self-administered and research-coordinator-administered digital assessments has generally found moderate to good agreement, with intraclass correlation coefficients ranging from approximately 0.59 to 0.83 across cognitive domains assessed, and no statistically significant difference in performance between self- and coordinator-administered conditions after accounting for testing order and device type.

Digital assessment platforms have also demonstrated clinical utility beyond the detection of age-related cognitive decline. Shibata et al. (2024) applied a remote digital cognitive assessment battery to detect cognitive deficits associated with hippocampal atrophy in patients with autoimmune limbic encephalitis, finding significant impairment across memory, visuospatial, executive, and language domains relative to healthy controls, with global digital assessment scores correlating significantly with established neuropsychological measures. In oncology populations, Root et al. (2023) established adequate test-retest reliability and internal validity for a remote, computerized neurocognitive battery designed to assess breast cancer- and treatment-related cognitive dysfunction. In low-resource settings, Paddick et al. (2021) demonstrated the feasibility and effectiveness of app-based cognitive screening administered by lay health workers for dementia detection in rural Tanzania, illustrating the potential for digital tools to extend screening capacity into settings historically underserved by specialist neuropsychological services. Collectively, this body of evidence suggests that digital and remote cognitive screening modalities are approaching, and in some domains exceeding, the diagnostic performance of traditional paper-based bedside instruments, while simultaneously expanding access to underserved populations; however, questions regarding longitudinal practice effects, equity of access to requisite technology, and standardization of digital normative data remain areas of active investigation.

Table 2. Comparative Diagnostic Accuracy of Selected Bedside Cognitive Screening Instruments

Instrument	Target Condition	Reported Cutoff	Sensitivity	Specificity	Key Source
MMSE	Dementia (community)	<24	85%	90%	Creavin et al. (2016)
MMSE	Dementia (community)	<25	87%	82%	Creavin et al. (2016)
MoCA	MCI	<23	73.5%	91.3%	Islam et al. (2023)
MoCA	MCI	<26 (original)	83.8%	70.8%	Islam et al. (2023)
Mini-Cog	Dementia (population-based)	Algorithm-based	76%	89%	Borson et al. (2003)
Mini-Cog	Dementia (primary care, low-bias study)	Algorithm-based	76%	73%	Seitz et al. (2021)
SLUMS	Major NCD	≤20	93–98%	96–100%	Spencer et al. (2022)
SLUMS	Mild NCD	≤26	74–95%	65–76%	Spencer et al. (2022)
RUDAS	Dementia (culturally diverse samples)	Variable	52–94%	70–98%	Chithiramohan et al. (2024)

Note. Reported values reflect pooled or representative estimates from the cited sources; substantial heterogeneity exists across populations, settings, and study quality, and individual studies within each meta-analysis or review may report divergent estimates. NCD = neurocognitive disorder.

Formal Neuropsychological Testing

Formal neuropsychological testing represents the second, more intensive tier of the cognitive assessment pathway, distinguished from bedside screening by its administration by licensed neuropsychologists, its considerably greater time investment (typically two to eight hours, occasionally distributed across multiple sessions), its reliance on extensively normed and standardized instruments, and its capacity to generate a detailed, domain-specific cognitive profile rather than a single composite score. Shaughnessy and Weintraub (2024), summarizing the Alzheimer's Association's Diagnostic Evaluation, Testing, Counseling and Disclosure (DETeCD-ADRD) clinical practice guideline, emphasize that neuropsychological evaluation supports early and accurate diagnosis and staging, characterization of the individual clinical profile, longitudinal tracking of trajectory, and the formulation of recommendations tailored to the specific needs of the patient and care team. Formal testing is generally indicated when bedside screening results are ambiguous or discordant with clinical presentation, when differential diagnosis among competing etiologies is required, when a detailed baseline is needed prior to initiation of disease-modifying therapy, or when functional capacity must be characterized for legal, occupational, or caregiving purposes.

1 Domains Assessed and Representative Instruments

Comprehensive neuropsychological assessment systematically evaluates multiple cognitive domains using standardized, psychometrically validated instruments with published age-, education-, and in some cases sex- and race-adjusted normative data (Lezak, Howieson, Bigler, & Tranel, 2012). Attention and processing speed are commonly assessed using measures such as the Trail Making Test Part A and Digit Symbol/Symbol Digit Modalities Test. Executive functioning is assessed through instruments including the Trail Making Test Part B, the Wisconsin Card Sorting Test, and the Stroop Color-Word Test; the Wisconsin Card Sorting Test in particular has demonstrated utility in characterizing executive dysfunction following traumatic brain injury (Gómez-de-Regil, 2020). Memory is evaluated through list-learning paradigms such as the California Verbal Learning Test and the Rey Auditory Verbal Learning Test, as well as through visual memory measures such as the Rey-Osterrieth Complex Figure Test, which additionally captures visuoconstructional and organizational abilities. Language is assessed through confrontation naming tasks such as the Boston Naming Test and through verbal fluency paradigms. General intellectual functioning is typically characterized using a Wechsler intelligence scale, and mood and personality functioning are frequently assessed using instruments such as the Beck Depression Inventory or the Minnesota Multiphasic Personality Inventory, given the well-documented potential for affective disorders to confound the interpretation of cognitive test performance (Lezak et al., 2012).

A defining methodological feature of formal neuropsychological assessment, relative to bedside screening, is the integration of standardized psychometric data with detailed clinical history, behavioural observation during testing, and collateral informant report. This integrative approach allows the examining neuropsychologist to establish an individualized expected-performance profile based on the patient's demographic characteristics and premorbid functioning, and subsequently to evaluate whether observed performance falls below that expected level in a pattern consistent with a specific neurological or psychiatric etiology.

2 Diagnostic Frameworks: Conventional versus Neuropsychological (Actuarial) Criteria for MCI

A substantial body of recent research has examined the comparative diagnostic performance of conventional (Petersen/Winblad) MCI criteria, which rely primarily on subjective cognitive complaint corroborated by a single below-threshold cognitive test score, against more comprehensive neuropsychological (actuarial, or Jak/Bondi) criteria, which require impairment across multiple measures within a cognitive domain or across multiple domains before a diagnosis of MCI is assigned (Jak et al., 2009). Kim, Kim, Edmonds, and Bondi (2024) directly compared these two diagnostic approaches within a clinical sample, finding 87.82% concordance between conventional and neuropsychological diagnoses overall, but identifying a clinically meaningful subgroup of individuals classified as impaired under conventional criteria who were reclassified as cognitively normal under the more stringent, multi-measure neuropsychological criteria. This discordant subgroup exhibited higher depression scores and, counterintuitively, better objective memory and executive function performance than those classified as impaired under both systems, suggesting that reliance on conventional single-test criteria alone may lead to overdiagnosis of MCI, particularly among individuals whose single below-threshold score reflects normal test variability, transient affective symptoms, or measurement error rather than genuine incipient neurodegeneration (Kim et al., 2024). These findings have direct implications for the interpretation of both bedside screening results, which are inherently limited to a small number of measures per domain, and formal neuropsychological batteries, which by virtue of their breadth are better positioned to apply actuarial, multi-measure diagnostic criteria and thereby reduce misclassification.

3 Performance Validity Testing

A methodological feature unique to formal neuropsychological assessment, largely absent from bedside screening protocols, is the systematic evaluation of performance validity—that is, the degree to which obtained test results credibly reflect the examinee's genuine cognitive capacity rather than volitional underperformance, symptom exaggeration, or insufficient engagement with testing. Sherman, Slick, and Iverson (2020), in a twenty-year update of the influential Malingered Neurocognitive Dysfunction criteria, articulate a multidimensional framework for classifying the credibility

of neuropsychological data, incorporating evidence from dedicated performance validity tests, atypical patterns of performance on standard clinical measures, and behavioral observations made during the evaluation. Performance validity assessment has assumed particular prominence in forensic, disability, and personal injury evaluation contexts, where the potential for secondary gain necessitates careful evaluation of response validity before test results can be considered diagnostically meaningful (Bosso et al., 2022; Panico, Fonzo, Verde, Lancia, & Trojano, 2025).

Contemporary research in this domain has emphasized both the refinement of existing performance validity instruments and the development of novel measures designed to resist coaching and detection avoidance. Panico et al. (2025) describe the development of the Test of Malingering in Abstraction Skills, designed to balance sufficient apparent difficulty to avoid overt detection as an effort measure against sufficient actual ease to remain insensitive to genuine cognitive impairment, illustrating the ongoing psychometric sophistication required in this subfield. Finley (2024) further discusses the emerging role of digital technology, including computerized adaptive testing and artificial intelligence-assisted analysis, in enhancing the sensitivity and efficiency of performance validity assessment, an area of active methodological development. The integration of performance validity testing into standard neuropsychological practice represents a critical safeguard against both false-positive and false-negative diagnostic conclusions and constitutes one of the most significant methodological distinctions between formal neuropsychological assessment and bedside cognitive screening, which generally lacks any analogous validity-checking mechanism.

4 Differential Diagnosis Across Dementia Subtypes

A principal clinical contribution of formal neuropsychological testing lies in its capacity to generate domain-specific cognitive profiles that assist in differentiating among the major neurodegenerative dementia syndromes, each of which is associated with a relatively characteristic, though not invariant, pattern of cognitive impairment. Alzheimer's disease dementia typically presents with early and prominent impairment of episodic memory, particularly delayed recall and recognition, accompanied by relatively later involvement of executive, language, and visuospatial domains as the disease progresses, consistent with the biologically staged model articulated by Jack et al. (2024). Dementia with Lewy bodies, whose diagnostic consensus criteria were most recently updated in the fourth report of the DLB Consortium (McKeith et al., 2017), is characterized by prominent deficits in attention, executive function, and visuospatial processing, often with relatively preserved episodic memory in earlier stages, accompanied by core clinical features including fluctuating cognition, recurrent visual hallucinations, rapid eye movement sleep behavior disorder, and spontaneous parkinsonism. Behavioural variant frontotemporal dementia, for which revised diagnostic criteria were established by Rascovsky et al. (2011), is characterized primarily by early and prominent

behavioural and personality change alongside executive dysfunction, with relative sparing of episodic memory and visuospatial function in earlier stages, distinguishing it neuropsychologically from typical Alzheimer's disease presentations. The primary progressive aphasia syndromes, classified according to the framework established by Gorno-Tempini et al. (2011), are distinguished by predominant, isolated language impairment as the presenting and most prominent feature, with further subclassification into semantic, nonfluent/agrammatic, and logopenic variants based on the specific pattern of linguistic breakdown. Vascular cognitive impairment, for which consensus classification criteria were articulated by Skrobot et al. (2017), typically presents with prominent executive dysfunction and processing speed impairment, often with a stepwise rather than gradually progressive course and a temporal relationship to cerebrovascular events, distinguishing it from the more insidious onset characteristic of primary neurodegenerative dementias.

Table 3. Characteristic Neuropsychological Profiles across Major Dementia Subtypes

Dementia Subtype	Diagnostic Framework	Predominant Early Cognitive Deficits	Relatively Spared Domains (Early Stage)	Key Reference
Alzheimer's disease	NIA-AA 2024 revised criteria	Episodic memory (encoding/delayed recall)	Executive function, language (until later stages)	Jack et al. (2024)
Dementia with Lewy bodies	DLB Consortium 4th consensus report	Attention, executive function, visuospatial processing	Episodic memory (relatively)	McKeith et al. (2017)
Behavioural variant frontotemporal dementia	Rascovsky revised criteria	Executive function, behaviour/personality change	Episodic memory, visuospatial function	Rascovsky et al. (2011)
Primary progressive aphasia (semantic, nonfluent, logopenic variants)	Gorno-Tempini classification	Language (variant-specific pattern)	Non-linguistic domains (until later stages)	Gorno-Tempini et al. (2011)
Vascular cognitive impairment	VICCCS consensus classification	Executive function, processing speed	Variable; often stepwise course	Skrobov et al. (2017)

Note. Profiles reflect characteristic, group-level patterns; substantial individual variability and mixed pathology are common in clinical practice. NIA-AA = National Institute on Aging–Alzheimer's Association; DLB = dementia with Lewy bodies; VICCCS = Vascular Impairment of Cognition Classification Consensus Study.

5 Formal Neuropsychological Testing in Special Populations

Formal neuropsychological assessment plays a particularly critical role in populations for whom bedside screening instruments are known to be insensitive or otherwise inadequate. In traumatic brain injury (TBI), neuropsychological evaluation has been characterized as the best available predictor of functional outcome, given its capacity to characterize the specific pattern and severity of cognitive sequelae across attention, processing speed, executive function, and memory domains (Torregrossa et al., 2023). Torregrossa et al. (2023), in a comprehensive review with clinical recommendations, note that general cognitive screening instruments such as the MMSE may be appropriately administered during the acute and subacute phases following TBI, while more specific instruments, including the Wisconsin Card Sorting Test and Trail Making Test, are recommended for detailed characterization of executive function during postacute and chronic phases, allowing clinicians to tailor rehabilitation strategies to the individual's specific cognitive profile. Repeat neuropsychological assessment is common in TBI rehabilitation contexts to track recovery trajectory, though clinicians must account for practice effects, which can artificially inflate scores upon repeat administration and are typically mitigated through the use of alternate, psychometrically equivalent test forms. Cluster-analytic approaches applied to large samples of patients with acquired brain injury undergoing cognitive rehabilitation have identified distinct neuropsychological subgroups characterized by differential performance profiles, informing more individualized approaches to web-based cognitive rehabilitation programming (García-Rudolph et al., 2021).

Formal testing is likewise essential in populations characterized by low educational attainment, limited literacy, or cultural and linguistic backgrounds divergent from those in which standard cognitive instruments were normed and validated, given the demonstrated propensity of screening instruments to misclassify such individuals (Chithiramohan et al., 2024). In such populations, formal neuropsychological evaluation, particularly when informed by culturally and linguistically appropriate normative data, allows for more individualized interpretation than a fixed screening cutoff can provide. Similarly, in psychiatric populations, where cognitive complaints may reflect a complex interplay of primary cognitive impairment, medication effects, and mood or psychotic symptomatology, formal testing that incorporates concurrent assessment of affective and psychotic symptoms, along with performance validity indices, provides diagnostic clarity that brief screening cannot achieve. Recent work examining working memory precision and associative binding among individuals with MCI illustrates the continuing refinement of specialized neuropsychological paradigms capable of detecting subtle cognitive alterations not captured by conventional screening or even standard neuropsychological batteries (Jia, Woltering, Deutz, et al., 2023).

Clinical Algorithm for Cognitive Assessment

Synthesizing the evidence reviewed above, a stepwise clinical algorithm can be constructed to guide clinicians through the cognitive assessment pathway, from initial identification of concern through formal diagnosis and, where appropriate, biomarker confirmation. Such an algorithm is broadly consistent with the approach recommended in the American Academy of Neurology practice guideline (Petersen et al., 2018) and the Alzheimer's Association DETeCD-ADRD clinical practice guideline (Shaughnessy & Weintraub, 2024), both of which emphasize a graduated approach beginning with case identification and proceeding through progressively more specialized levels of evaluation as clinically indicated.

Table 4. Stepwise Clinical Algorithm for Cognitive Assessment

Step	Clinical Action	Instruments/Methods	Decision Point
1. Case identification	Elicit subjective cognitive concern from patient, informant, or clinical observation; review health risk assessment	Clinical interview; standardized health risk assessment tools	If concern present, proceed to Step 2; if absent, routine monitoring
2. Delirium/acute confounder exclusion (if applicable acute setting)	Rule out delirium or acute confounding medical condition before interpreting cognitive test results	Confusion Assessment Method (CAM); 4AT; medical workup	If delirium present, treat underlying cause before further cognitive testing; if absent, proceed to Step 3
3. Bedside cognitive screening	Administer brief, validated screening instrument appropriate to setting, education, and cultural/linguistic background	MMSE, MoCA, Mini-Cog, SLUMS, or culture-fair alternative (e.g., RUDAS)	Score within normal range: routine monitoring/reassessment interval; score suggestive of impairment: proceed to Step 4
4. Functional assessment	Evaluate impact of cognitive symptoms on instrumental and basic activities of daily living	Structured functional/IADL questionnaires; informant interview	Functional independence preserved: consistent with mild NCD/MCI; functional impairment present: consistent with major NCD
5. Medical and laboratory workup	Exclude reversible or contributory medical causes (e.g., thyroid dysfunction, vitamin B12 deficiency, depression, medication effects)	Laboratory studies; mood screening; medication review; structural neuroimaging	Reversible cause identified and treated: reassess cognition after treatment; no reversible cause: proceed to Step 6
6. Referral for formal neuropsychological testing	Refer when screening results are ambiguous, discordant with function, or detailed differential diagnosis is required	Comprehensive neuropsychological battery with performance validity testing	Profile consistent with specific etiology: proceed to targeted biomarker workup; diagnosis unclear: consider longitudinal reassessment
7. Biomarker and etiological workup	Confirm underlying biological etiology, particularly prior to consideration of disease-modifying therapy	Amyloid/tau PET, CSF biomarkers, blood-based biomarkers, structural/functional neuroimaging	Biomarker-confirmed Alzheimer's pathology with clinical impairment: eligible for staged diagnosis and treatment planning per NIA-AA criteria
8. Diagnosis, staging, and disclosure	Integrate clinical, cognitive, functional, and biomarker data into a formal diagnosis and staging	DSM-5-TR criteria; NIA-AA 2024 biological-clinical staging matrix	Diagnosis disclosed with counseling; care/treatment plan established

Note. Algorithm synthesized from recommendations in Petersen et al. (2018), Shaughnessy and Weintraub (2024), and Jack et al. (2024). IADL = instrumental activities of daily living; PET = positron emission tomography; CSF = cerebrospinal

Comparative Analysis: Bedside Screening versus Formal Neuropsychological Testing

Although bedside screening and formal neuropsychological testing are frequently discussed in tandem, they differ fundamentally across nearly every methodological and clinical dimension, and an appreciation of these differences is essential to appropriate instrument selection. Bedside screening instruments are optimized for speed, accessibility, and

broad applicability across non-specialist clinical settings, at the cost of diagnostic granularity and, frequently, specificity. Formal neuropsychological testing sacrifices brevity and accessibility in exchange for diagnostic precision, domain-specific characterization, integration of performance validity assessment, and the capacity to support differential diagnosis among etiologically distinct conditions that may present with superficially similar screening results. Table 5 summarizes these distinctions across key dimensions of clinical relevance.

Table 5. Comparative Summary: Bedside Cognitive Screening versus Formal Neuropsychological Testing

Dimension	Bedside Screening	Formal Neuropsychological Testing
Typical duration	5–15 minutes	2–8+ hours, potentially multiple sessions
Administering clinician	Physicians, nurses, medical assistants, trained lay health workers	Licensed neuropsychologists (doctoral-level training)
Primary purpose	Case detection, triage, longitudinal monitoring	Diagnosis, differential diagnosis, domain-specific profiling, capacity evaluation
Output	Single (or few) composite score(s)	Multidimensional profile across cognitive, behavioural, and emotional domains
Sensitivity to MCI	Variable; generally moderate, improved with lowered cutoffs	High, particularly with actuarial/multi-measure criteria
Specificity	Variable; often reduced when sensitivity maximized	Generally higher due to comprehensive, multi-measure assessment
Performance validity assessment	Not routinely incorporated	Integral component of standard practice
Cultural/educational adaptability	Limited unless culture-fair instrument specifically selected	Higher, given individualized normative comparison and clinical judgment
Accessibility/cost	Low cost; widely accessible, including via digital/remote platforms	Higher cost; limited by neuropsychologist availability, particularly in underserved regions
Role in biomarker-era diagnosis (e.g., anti-amyloid therapy eligibility)	Supports initial identification of candidates for further workup	Provides detailed functional-cognitive staging required alongside biomarker confirmation

Note. Synthesized from evidence reviewed in Sections 3 and 4, including Petersen et al. (2018), Lezak et al. (2012), and Shaughnessy and Weintraub (2024).

This comparative analysis reinforces the conclusion that screening and formal testing serve complementary rather than competing functions. Screening instruments efficiently triage large populations, directing limited specialist neuropsychological resources toward those individuals most likely to benefit from comprehensive evaluation, while formal testing provides the diagnostic precision necessary for individualized clinical decision-making, particularly in cases of diagnostic ambiguity, atypical presentation, or high-stakes treatment decisions such as eligibility for disease-modifying anti-amyloid therapy.

Emerging Directions in Cognitive Testing

The field of cognitive testing is undergoing rapid transformation, driven by three converging developments: the maturation of digital and remote assessment technologies, the emergence of blood-based biomarkers for Alzheimer's disease pathology, and the increasing application of artificial intelligence and machine learning methods to cognitive test interpretation.

The proliferation of digital cognitive assessment platforms, reviewed extensively in Section 3.7, is increasingly extending beyond simple digitization of traditional paper-based instruments toward the development of novel digital biomarkers with potentially superior sensitivity to preclinical and prodromal disease states. Thompson et al. (2023) demonstrated that certain digital cognitive tasks can achieve discriminative accuracy for amyloid status exceeding that of conventional bedside screening, while ecological momentary assessment and ambulatory, smartphone-based monitoring approaches allow for repeated, longitudinal characterization of cognitive fluctuation in naturalistic settings, a capability entirely absent from single-timepoint bedside or laboratory-based assessment. Continued methodological development is needed to establish robust normative datasets for digital instruments across diverse populations, to characterize practice effects arising from repeated remote administration, and to ensure equitable access to the technological infrastructure such assessments require, particularly among older adults and individuals in low-resource settings (Paddick et al., 2021).

The introduction of blood-based biomarkers into the revised NIA-AA diagnostic criteria (Jack et al., 2024) represents a further inflection point in the cognitive assessment pathway. Whereas biomarker confirmation of Alzheimer's pathology has historically required either positron emission tomography imaging or cerebrospinal fluid analysis via lumbar puncture, both of which are costly, invasive, or of limited availability outside specialized centers, blood-based biomarkers offer the prospect of substantially more accessible and scalable biological confirmation. Jack et al. (2024/2025), discussing the application of the revised criteria in drug development and clinical practice, note that improving the availability and accuracy of plasma biomarkers is expected to substantially enhance the

practical applicability of biologically anchored diagnostic criteria, particularly as disease-modifying therapies such as lecanemab (van Dyck et al., 2023) and donanemab (Sims et al., 2023) become more widely available and require biomarker-confirmed diagnosis prior to treatment initiation. The integration of blood-based biomarkers with cognitive testing data, whether obtained through bedside screening or formal neuropsychological assessment, will likely become an increasingly central feature of the diagnostic pathway over the coming years, potentially altering the sequencing and relative emphasis placed on cognitive versus biological data in future iterations of clinical practice guidelines.

Finally, artificial intelligence and machine learning methods are increasingly being applied both to enhance the diagnostic yield of existing cognitive instruments and to develop novel assessment paradigms. Kim et al. (2025) demonstrated that machine learning models incorporating MMSE data alongside additional cognitive measures achieve meaningfully improved diagnostic accuracy relative to MMSE scores considered in isolation, illustrating the potential for computational methods to extract additional diagnostic value from existing, widely available screening data without requiring entirely new instruments. Similarly, Finley (2024) discusses the emerging application of artificial intelligence to performance validity assessment within formal neuropsychological testing, suggesting that computational methods may eventually assist neuropsychologists in detecting subtle, non-credible response patterns not readily apparent through conventional clinical judgment alone. These developments collectively suggest a diagnostic future in which cognitive testing, whether at the bedside or within formal neuropsychological evaluation, is increasingly integrated with digital, biomarker, and computational data streams, potentially yielding diagnostic precision substantially exceeding that achievable through any single assessment modality in isolation.

Limitations and Methodological Considerations

Notwithstanding the substantial evidence base reviewed in this paper, several methodological limitations warrant explicit consideration by clinicians and researchers alike. First, the diagnostic accuracy literature for both bedside screening and formal neuropsychological instruments is characterized by considerable heterogeneity in study design, reference standards, population characteristics, and disease prevalence across settings, complicating direct comparison of sensitivity and specificity estimates across studies. The Cochrane reviews examining the MMSE (Creavin et al., 2016; Arevalo-Rodriguez et al., 2021) and the Mini-Cog (Seitz et al., 2021) each identified substantial methodological limitations, including high or unclear risk of bias in patient selection, across the majority of included primary studies, urging caution in the uncritical application of pooled accuracy estimates to individual clinical decision-making.

Second, the pervasive influence of educational attainment, literacy, language proficiency, and cultural background on cognitive test performance constitutes a persistent threat to the validity of cognitive assessment across both bedside screening and, to a lesser but still meaningful extent, formal neuropsychological testing. As documented extensively in Section 3.5, standard cutoffs derived from predominantly Western, highly educated validation samples frequently generalize poorly to populations with limited formal education, non-native language proficiency, or divergent cultural reference points, elevating risk of both false-positive and false-negative classification (Chithiramohan et al., 2024). While formal neuropsychological testing partially mitigates this concern through individualized normative comparison and clinical judgment, the underlying normative datasets themselves frequently suffer from similar demographic limitations, underscoring the ongoing need for more diverse and representative normative data development across all tiers of cognitive assessment.

Third, confounding by acute or subacute medical, affective, and pharmacological factors represents a persistent challenge to accurate cognitive assessment interpretation, irrespective of the tier of testing employed. Delirium, in particular, can profoundly distort cognitive test performance and, if unrecognized, may lead to inappropriate diagnosis of a chronic neurocognitive disorder in a patient experiencing a transient, reversible confusional state; the CAM and related delirium screening instruments (Inouye et al., 1990; Muser et al., 2022) exist precisely to guard against this form of misclassification, though implementation research indicates that consistent, protocol-adherent delirium screening remains incompletely achieved even in well-resourced hospital systems (Penfold et al., 2024). Similarly, depressive symptomatology can produce a pattern of cognitive underperformance, sometimes termed pseudodementia, that may be difficult to distinguish from genuine neurodegenerative impairment without careful integration of affective symptom assessment, as illustrated by the Kim et al. (2024) finding that individuals reclassified from impaired to normal under stringent neuropsychological criteria exhibited elevated depression scores.

Finally, the rapid pace of technological change in digital and biomarker-based assessment, while offering considerable promise, also introduces new sources of uncertainty, including limited long-term validation data, evolving normative standards, and unresolved questions regarding equity of access across socioeconomic and geographic strata (Paddick et al., 2021). Clinicians and health systems adopting these emerging technologies should do so with appropriate awareness of their still-maturing evidence base, alongside the well-established, extensively validated traditional instruments reviewed throughout this paper.

Discussion

The evidence reviewed in this paper collectively supports a stepped, context-sensitive approach to cognitive assessment in which bedside screening and formal neuropsychological testing occupy distinct, complementary, and non-interchangeable roles. Bedside instruments such as the MMSE, MoCA, Mini-Cog, and SLUMS provide efficient, broadly accessible means of identifying individuals who may warrant further evaluation, and their continued refinement, through revised cutoffs (Islam et al., 2023), culture-fair adaptation (Chithiramohan et al., 2024), and digital delivery (Kochan et al., 2022; Morrissey et al., 2024), has substantially expanded their clinical utility since their original development. At the same time, the accumulated evidence base makes clear that no bedside instrument, however well validated, can substitute for the diagnostic precision, domain-specific characterization, and performance validity safeguards afforded by formal neuropsychological testing, particularly in cases involving diagnostic ambiguity, atypical presentation, differential diagnosis among competing dementia etiologies, or high-stakes treatment eligibility determinations in the era of disease-modifying anti-amyloid therapy (van Dyck et al., 2023; Sims et al., 2023).

The integration of biologically anchored diagnostic frameworks, most recently articulated in the 2024 NIA-AA revised criteria (Jack et al., 2024), further underscores the continuing centrality of cognitive testing within a broader, increasingly multimodal diagnostic paradigm. Biomarker positivity alone does not constitute a clinical diagnosis of dementia or MCI in the functional sense embodied by both the DSM-5-TR (American Psychiatric Association, 2022) and prevailing clinical practice; rather, cognitive and functional assessment, whether through validated screening or comprehensive neuropsychological evaluation, remains indispensable to establishing the clinical-syndromic dimension of diagnosis that biomarkers alone cannot provide (Shaughnessy & Weintraub, 2024). This complementarity suggests that, rather than being rendered obsolete by advances in biomarker science, cognitive testing is likely to become increasingly integrated with biological data, with digital cognitive assessment in particular positioned to serve as a bridge between traditional bedside screening and biomarker-confirmed diagnosis, given its capacity for repeated, longitudinal, and potentially more sensitive characterization of subtle cognitive change (Thompson et al., 2023).

From a health systems perspective, the evidence also highlights persistent challenges in equitable implementation. The demonstrated vulnerability of standard cognitive screening instruments to educational, linguistic, and cultural bias (Chithiramohan et al., 2024), combined with the limited availability of formal neuropsychological services in many regions and healthcare systems, risks perpetuating disparities in the timely and accurate diagnosis of cognitive impairment

across demographic groups. The expansion of culture-fair screening instruments, lay-health-worker-administered digital tools (Paddick et al., 2021), and telehealth-delivered neuropsychological services (Owens et al., 2020) each represent partial responses to this challenge, though continued investment in the development, validation, and equitable deployment of such approaches remains necessary.

Conclusion

Cognitive testing spans a continuum from rapid, broadly accessible bedside screening to comprehensive, specialist-administered formal neuropsychological evaluation, with each tier fulfilling a distinct and indispensable function within the overall diagnostic pathway for neurocognitive disorders. The MMSE and MoCA remain foundational instruments for initial detection and longitudinal monitoring of cognitive impairment, notwithstanding well-documented limitations related to educational, cultural, and linguistic bias that have prompted the development of alternative and culture-fair instruments such as the SLUMS and RUDAS. Formal neuropsychological testing offers substantially greater diagnostic precision, domain-specific profiling, and performance validity assurance, positioning it as the essential second tier of assessment for cases requiring differential diagnosis, detailed functional characterization, or high-stakes treatment decision-making. The integration of biologically anchored diagnostic frameworks, exemplified by the 2024 NIA-AA revised criteria, alongside the rapid maturation of digital cognitive assessment technologies and blood-based biomarkers, is actively reshaping the cognitive assessment landscape, without diminishing the fundamental necessity of rigorous cognitive and functional evaluation. A stepped, context-sensitive approach, employing validated screening tools to identify individuals warranting further assessment, followed by formal neuropsychological evaluation and, where indicated, biomarker confirmation, continues to represent best practice in the evaluation of cognitive function across clinical settings.

References

- Abdelnour, C. (2025). The Alzheimer's Association revised criteria for diagnosis and staging of Alzheimer's disease—Where does Lewy body dementia stand? *Alzheimer's & Dementia*, 21(Suppl. 1). <https://doi.org/10.1002/alz.089685>
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroup. *Alzheimer's & Dementia*, 7(3), 270–279.
- Alzheimer's Association. (2024). 2024 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 20(5), 3708–4002.
- American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). American Psychiatric Publishing.
- Arevalo-Rodriguez, I., Smailagic, N., Roqué-Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Bonfill Cosp, X., & Cullum, S. (2021). Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, 2021(7), CD010783. <https://doi.org/10.1002/14651858.CD010783.pub3>
- Borson, S., Scanlan, J. M., Chen, P., & Ganguli, M. (2003). The Mini-Cog as a screen for dementia: Validation in a population-based sample. *Journal of the American Geriatrics Society*, 51(10), 1451–1454.
- Bosso, T., Vischia, F., Keller, R., Vai, D., Imperiale, D., & Vercelli, A. (2022). A case report and literature review of cognitive malingering and psychopathology. *Frontiers in Psychiatry*, 13, Article 981475. <https://doi.org/10.3389/fpsyt.2022.981475>
- Chen, Y.-X., Liang, N., Li, X.-L., Yang, S.-H., Wang, Y.-P., & Shi, N.-N. (2021). Diagnosis and treatment for mild cognitive impairment: A systematic review of clinical practice guidelines and consensus statements. *Frontiers in Neurology*, 12, Article 719849. <https://doi.org/10.3389/fneur.2021.719849>
- Chithiramohan, T., Santhosh, S., Threlfall, G., Hull, L., Mukaetova-Ladinska, E., Subramaniam, H., & Beishon, L. (2024). Culture-fair cognitive screening tools for assessment of cognitive impairment: A systematic review. *Alzheimer's & Dementia Reports*, 8, Article ADR-230194.
- Cornea, A., Simu, M., & Rosca, E. C. (2022). Montreal Cognitive Assessment for evaluating cognitive impairment in subarachnoid hemorrhage: A systematic review. *Journal of Clinical Medicine*, 11(16), Article 4679. <https://doi.org/10.3390/jcm11164679>
- Creavin, S. T., Wisniewski, S., Noel-Storr, A. H., Trevelyan, C. M., Hampton, T., Rayment, D., Thom, V. M., Nash, K. J. E., Elhamoui, H., Milligan, R., Patel, A. S., Tsivos, D. V., Wing, T., Phillips, E., Kellman, S. M., Shackleton, H. L., Singleton, G. F., Neale, B. E., Watton, M. E., & Cullum, S. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews*, 2016(1), CD011145.
- Finley, J.-C. A. (2024). Performance validity testing: The need for digital technology and where to go from here. *Frontiers in Psychology*, 15, Article 1452462. <https://doi.org/10.3389/fpsyg.2024.1452462>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- García-Rudolph, A., García-Molina, A., Opisso, E., Tormos, J. M., Madai, V. I., Frey, D., & Bernabeu, M. (2021). Neuropsychological assessments of patients with acquired brain injury: A cluster analysis approach to address heterogeneity in web-based cognitive rehabilitation. *Frontiers in Neurology*, 12, Article 701946. <https://doi.org/10.3389/fneur.2021.701946>

- Gómez-de-Regil, L. (2020). Assessment of executive function in patients with traumatic brain injury with the Wisconsin Card-Sorting Test. *Brain Sciences*, 10(10), Article 699. <https://doi.org/10.3390/brainsci10100699>
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76(11), 1006–1014.
- Huynh, D., Patterson, M., & Huang, B. (2025). Concurrent validity of BrainCheck compared to SLUMS in older adults using real-world clinical data. *Innovation in Aging*, 9(Suppl. 1). <https://doi.org/10.1093/geroni/igaf122.4090>
- Inouye, S. K., van Dyck, C. H., Alessi, C. A., Balkin, S., Siegel, A. P., & Horwitz, R. I. (1990). Clarifying confusion: The confusion assessment method. *Annals of Internal Medicine*, 113(12), 941–948.
- Islam, M., Iqbal, U., Walther, B. A., Nguyen, P.-A., Yang, H.-C., Poly, T. N., Huang, C.-W., Jian, W.-S., & Li, Y.-C. (2023). Accuracy of the Montreal Cognitive Assessment tool for detecting mild cognitive impairment: A systematic review and meta-analysis. *Alzheimer's & Dementia*, 19(8), 3436–3447. <https://doi.org/10.1002/alz.13040>
- Jack, C. R., Jr., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., Molinuevo, J. L., Okonkwo, O. C., Pani, L., Rafii, M. S., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Teunissen, C. E., & Carrillo, M. C. (2024). Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, 20(8), 5143–5169. <https://doi.org/10.1002/alz.13859>
- Jack, C. R., Jr., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Jagust, W., McDade, E., Molinuevo, J. L., Okonkwo, O. C., Pani, L., Rafii, M. S., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Teunissen, C. E., Weber, C. J., & Carrillo, M. C. (2025). Application of the revised criteria for diagnosis and staging of Alzheimer's disease: Drug development and clinical practice. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 11(1), Article e70013. <https://doi.org/10.1002/trc2.70013>
- Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *American Journal of Geriatric Psychiatry*, 17(5), 368–375.
- Jia, Y., Woltering, S., Deutz, N., & colleagues. (2023). Working memory precision and associative binding in mild cognitive impairment. *Experimental Aging Research*, 49, 1–19.
- Kalbe, E., Folkerts, A.-K., Witt, K., Buhmann, C., & Liepelt-Scarfone, I. (2024). German Society of Neurology guidelines for the diagnosis and treatment of cognitive impairment and affective disorders in people with Parkinson's disease: New spotlights on diagnostic procedures and non-pharmacological interventions. *Journal of Neurology*, 271, 6160–6181. <https://doi.org/10.1007/s00415-024-12503-0>
- Kim, H., Kim, Y., Edmonds, E. C., & Bondi, M. W. (2024). Comparison of cognitive impairment diagnosis criteria in clinical settings: Conventional vs. neuropsychological. *Alpha Psychiatry*, 25(2), 212–219.
- Kim, H. G., et al. (2025). Dementia overdiagnosis in younger, higher educated individuals based on MMSE alone: Analysis using deep learning technology. *Journal of Korean Medical Science*, 40(9), Article e20. <https://doi.org/10.3346/jkms.2025.40.e20>
- Kochan, N. A., Heffernan, M., Valenzuela, M., Sachdev, P. S., Lam, B. C. P., Fiatarone Singh, M., & colleagues. (2022). Reliability, validity, and user-experience of remote unsupervised computerized neuropsychological assessments in community-living 55- to 75-year-olds. *Journal of Alzheimer's Disease*, 90(4), 1629–1645. <https://doi.org/10.3233/JAD-220665>
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.
- Loewenstein, D. A., Curiel, R. E., Duara, R., & Buschke, H. (2018). Novel cognitive paradigms for the detection of memory impairment in preclinical Alzheimer's disease. *Assessment*, 25(3), 348–359.
- McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J.-P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C. G., Bayston, A., Beach, T. G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., ... Kosaka, K. (2017). Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*, 89(1), 88–100.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroup. *Alzheimer's & Dementia*, 7(3), 263–269.
- Morrissey, S., Gillings, R., & Hornberger, M. (2024). Feasibility and reliability of online vs in-person cognitive testing in healthy older people. *PLOS ONE*, 19(8), Article e0309006. <https://doi.org/10.1371/journal.pone.0309006>
- Muser, O., Seiler, K., Bachnick, S., & colleagues. (2022). Delirium detection in hospitalized adults: The performance of the 4 'A's Test and the modified Confusion Assessment Method for the Emergency Department. A comparison study. *Bulletin of the National Research Centre*, 46, Article 179. <https://doi.org/10.1186/s42269-022-00863-4>
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699.

- Owens, A. P., Ballard, C., Beigi, M., Kalafatis, C., Brooker, H., Lavelle, G., Brown, R. G., & Zamboni, G. (2020). Implementing remote memory clinics to enhance clinical care during and after COVID-19. *Frontiers in Psychiatry*, 11, Article 579934.
- Paddick, S.-M., Yoseph, M., Gray, W. K., Ola, B., Dotchin, C., Chaya, E., Kisoli, A., Longdon, A., & Walker, R. (2021). Effectiveness of app-based cognitive screening for dementia by lay health workers in low resource settings: A validation and feasibility study in rural Tanzania. *Journal of Geriatric Psychiatry and Neurology*, 34(6), 613–621.
- Panico, F., Fonzo, E., Verde, A., Lancia, S., & Trojano, L. (2025). Proposal for a new tool assessing validity performance in forensic neuropsychological testing: The Test of Malingering in Abstraction Skills (TOMAS). *Neurological Sciences*. <https://doi.org/10.1007/s10072-025-08061-6>
- Penfold, C., & colleagues. (2024). Delirium detection tools show varying completion rates and positive score rates when used at scale in routine practice in general hospital settings: A systematic review. *Journal of the American Geriatrics Society*, 72(6), 1885–1901. <https://doi.org/10.1111/jgs.18751>
- Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., Gronseth, G. S., Marson, D., Pringsheim, T., Day, G. S., Sager, M., Stevens, J., & Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment. Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*, 90(3), 126–135.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303–308.
- Rajan, K. B., Weuve, J., Barnes, L. L., McAninch, E. A., Wilson, R. S., & Evans, D. A. (2021). Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). *Alzheimer's & Dementia*, 17(12), 1966–1975.
- Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini, M. L., Rosen, H., ... Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, 134(9), 2456–2477.
- Root, J. C., Gaynor, A. M., Ahsan, A., Jung, D., Schofield, E., Ryan, E., Li, Y., & Ahles, T. A. (2023). Remote, computerised cognitive assessment for breast cancer and treatment-related cognitive dysfunction: Psychometric characteristics of the Cogsuite neurocognitive battery. *Archives of Clinical Neuropsychology*, 38(5), 699–713. <https://doi.org/10.1093/arclin/acac111>
- Seitz, D. P., Chan, C. C., Newton, H. T., Gill, S. S., Herrmann, N., Smailagic, N., Nikolaou, V., & Fage, B. A. (2021). Mini-Cog for the detection of dementia within a primary care setting. *Cochrane Database of Systematic Reviews*, 2021(7), CD011415. <https://doi.org/10.1002/14651858.CD011415.pub3>
- Shaughnessy, L. W., & Weintraub, S. (2024). The role of neuropsychological assessment in the evaluation of patients with cognitive-behavioral change due to suspected Alzheimer's disease and other causes of cognitive impairment and dementia. *Alzheimer's & Dementia*, 21(1), Article e14363. <https://doi.org/10.1002/alz.14363>
- Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, 35(6), 735–764.
- Shibata, K., Attaallah, B., Tai, X.-Y., Trender, W., Hellyer, P. J., Hampshire, A., Irani, S. R., Manohar, S. G., & Husain, M. (2024). Remote digital cognitive assessment reveals cognitive deficits related to hippocampal atrophy in autoimmune limbic encephalitis: A cross-sectional validation study. *eClinicalMedicine*, 69, Article 102437. <https://doi.org/10.1016/j.eclinm.2024.102437>
- Sims, J. R., Zimmer, J. A., Evans, C. D., Lu, M., Ardayfio, P., Sparks, J., & colleagues. (2023). Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*, 330(6), 512–527.
- Skrobot, O. A., O'Brien, J., Black, S., Chen, C., DeCarli, C., Erkinjuntti, T., Ford, G. A., Kalaria, R. N., Pantoni, L., Pasquier, F., Roman, G. C., Wallin, A., Sachdev, P., Skoog, I., & Kehoe, P. G. (2017). The Vascular Impairment of Cognition Classification Consensus Study. *Alzheimer's & Dementia*, 13(6), 624–633.
- Song, W.-X., Wu, W.-W., Zhao, Y.-Y., Xu, H.-L., Chen, G.-C., Jin, S.-Y., & colleagues. (2023). Evidence from a meta-analysis and systematic review reveals the global prevalence of mild cognitive impairment. *Frontiers in Aging Neuroscience*, 15, Article 1227112.
- Spencer, R. J., Noyes, E. T., Bair, J. L., & Ransom, M. T. (2022). Systematic review of the psychometric properties of the Saint Louis University Mental Status (SLUMS) examination. *Clinical Gerontologist*, 45(3), 454–466. <https://doi.org/10.1080/07317115.2022.2032523>
- Thompson, L. I., & colleagues. (2023). Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(4), Article e12500. <https://doi.org/10.1002/dad2.12500>
- Torregrossa, W., Torrisi, M., De Luca, R., Casella, C., Rifci, C., Bonanno, M., & Calabrò, R. S. (2023). Neuropsychological assessment in patients with traumatic brain injury: A comprehensive review with clinical recommendations. *Biomedicine*, 11(7), Article 1991. <https://doi.org/10.3390/biomedicine11071991>
- van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., & colleagues. (2023). Lecanemab in early Alzheimer's disease. *New England Journal of Medicine*, 388(1), 9–21.

Neutropenia in Infants: A Comprehensive Narrative Review

Khawla Faraq Alzaitoni

Correspondence:

Dr Khawla Faraq Alzaitoni

Consultant pediatrician at Primary health care corporation (PHCC)

Doha, Qatar

Email: khawla_z82@yahoo.com

Received: August 2026. Accepted: September 2026; Published: October 2026.

Citation: Khawla Faraq Alzaitoni. Neutropenia in Infants: A Comprehensive Narrative Review. World Family Medicine.

October 2026; 24(7): 55 - 65 DOI: 10.5742/MEWFM.2026.241796

Abstract

Neutropenia in infants represents a heterogeneous group of disorders characterized by a reduced absolute neutrophil count (ANC), predisposing affected individuals to recurrent and potentially life-threatening bacterial and fungal infections. This narrative review provides a comprehensive synthesis of the current evidence regarding the epidemiology, etiology, classification, pathophysiology, clinical manifestations, diagnostic approach, differential diagnosis, management, prognosis, and future perspectives of neutropenia in the neonatal and infant population. Neutropenia may arise from transient physiological processes, immune-mediated mechanisms (alloimmune and autoimmune), congenital bone marrow failure syndromes (including severe congenital neutropenia, cyclic neutropenia, and Shwachman-Diamond syndrome), infections, drug-induced causes, or nutritional deficiencies. The diagnostic approach requires systematic evaluation integrating clinical history, serial complete blood counts, peripheral blood smear, antineutrophil antibody testing, bone marrow examination, and increasingly, next-generation sequencing for germline and somatic genetic variants. Management is tailored to the underlying etiology and severity, with recombinant human granulocyte colony-stimulating factor (rhG-CSF) serving as the cornerstone of therapy for severe and symptomatic forms. Prognosis varies markedly by etiology; while autoimmune neutropenia of infancy typically resolves spontaneously within 2–4 years, severe congenital neutropenia

carries a substantial risk of leukemic transformation. Emerging therapies, including targeted gene editing approaches, hold promise for definitive treatment of monogenic forms. This review consolidates contemporary knowledge to guide clinicians in the recognition, evaluation, and evidence-based management of neutropenia in infants.

Keywords: Neutropenia; Infant; Newborn; Severe Congenital Neutropenia; Autoimmune Neutropenia; Alloimmune Neutropenia; G-CSF; ELANE; Bone Marrow Failure

Introduction

Neutropenia, defined as an absolute neutrophil count (ANC) below the age-appropriate reference range, constitutes one of the most commonly encountered hematologic abnormalities in neonates and infants (1). The clinical significance of neutropenia lies in its association with increased susceptibility to bacterial and fungal infections, which can range from mild recurrent skin and mucosal infections to overwhelming sepsis with significant morbidity and mortality (2).

In infants, the interpretation of neutrophil counts is particularly challenging because reference values vary substantially according to gestational age, postnatal age, birth weight, ethnicity, altitude, and the source of the blood sample (capillary versus venous) (3,4).

The etiologic spectrum of infant neutropenia is remarkably broad, encompassing transient physiologic phenomena, immune-mediated destruction of neutrophils, inherited bone marrow failure syndromes, infections, drug-induced marrow suppression, and nutritional deficiencies (5).

Distinguishing among these causes is essential because management strategies, prognosis, and genetic counseling implications differ profoundly. For example, neonatal alloimmune neutropenia, caused by transplacental passage of maternal antineutrophil antibodies, is typically self-limited, whereas severe congenital neutropenia (SCN) due to ELANE mutations requires lifelong therapy and carries a substantial risk of malignant transformation (6,7).

Over the past two decades, advances in molecular genetics have transformed the understanding of inherited neutropenia syndromes.

The identification of causative genes, including ELANE, HAX1, G6PC3, CSF3R, WAS, GFI1, SBDS, and SRP54 has clarified the pathophysiology of these disorders and enabled targeted diagnostic testing (8,9). Simultaneously, the introduction of recombinant human granulocyte colony-stimulating factor (rhG-CSF) in the late 1980s revolutionized the management of severe neutropenia, dramatically reducing infection-related mortality (10). More recently, CRISPR-based gene editing technologies have emerged as potential curative strategies for monogenic neutropenia syndromes, with promising preclinical data published in 2024–2025 (11,12).

This narrative review aims to synthesize the contemporary evidence regarding neutropenia in infants, with particular emphasis on the most recent international guidelines, advances in molecular diagnostics, evidence-based management strategies, and emerging therapeutic frontiers. The review is intended for clinicians, including neonatologists, pediatric hematologists, infectious disease specialists, and general pediatricians, who encounter infants with neutropenia in diverse clinical settings.

Epidemiology

1 General Neonatal Neutropenia

Neutropenia is detected in approximately 1–5% of all newborns when broad ANC thresholds are applied, with substantially higher rates among preterm and low-birth-weight infants (13). In a prospective study of more than 1,000 neonates admitted to a tertiary neonatal intensive care unit, neutropenia ($\text{ANC} < 1.5 \times 10^9/\text{L}$) was identified in up to 8% of preterm infants compared with less than 1% of term infants (14). The higher prevalence in preterm infants reflects the immaturity of myeloid progenitor cell pools and the greater susceptibility of preterm neonates to perinatal complications that suppress neutrophil production.

2 Alloimmune Neonatal Neutropenia

Alloimmune neonatal neutropenia (ANN), analogous to hemolytic disease of the newborn, occurs when maternal antineutrophil antibodies cross the placenta and destroy fetal neutrophils. The estimated incidence ranges from 0.1% to 0.9% of live births, depending on the population and the diagnostic criteria applied (15,16). A Brazilian cross-sectional study of 10,000 neonates identified an incidence of 0.9%, with maternal-fetal incompatibility for HNA-1 (human neutrophil antigen-1) detected in 44.3% of cases and for HNA-3 in 14.8% (17). The most commonly implicated antigen is HNA-1a, although HNA-1b, HNA-1c, and HNA-3b have also been documented.

3 Autoimmune Neutropenia of Infancy

Primary autoimmune neutropenia of infancy (AIN) is among the most common causes of chronic neutropenia in children under 3 years of age, with an estimated incidence of 1 in 100,000 to 1 in 200,000 per year (18). The median age at diagnosis is 7–12 months, and the condition accounts for the majority of cases of chronic benign neutropenia in early childhood. Unlike ANN, AIN involves autoantibodies directed against the infant's own neutrophils and typically resolves spontaneously within 2–4 years.

4 Severe Congenital Neutropenia

Severe congenital neutropenia (SCN) is a rare disorder with an estimated incidence of 1–2 cases per million live births (19). The most common genetic form, ELANE-related SCN, accounts for approximately 50–60% of cases, followed by HAX1-related (Kostmann disease), G6PC3-related, and other rarer forms (20). Although rare, SCN is clinically significant because of its severe infection risk and the substantial lifetime risk of progression to myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML).

5 Ethnic and Geographic Variation

ANC reference ranges vary by ethnicity. Studies have demonstrated that infants of African descent have lower mean neutrophil counts compared with those of European descent, with some authors proposing ethnicity-specific reference ranges to avoid over-diagnosis of neutropenia (21). Geographic factors, including altitude, also influence neutrophil counts, with higher values observed at high altitudes due to chronic hypoxia-induced marrow stimulation (22).

Etiology and Classification

Neutropenia in infants can be classified along several axes: by duration (acute versus chronic), by mechanism (production defect versus increased destruction or sequestration), by inheritance pattern (congenital versus acquired), and by clinical severity. The most clinically useful classification stratifies cases into the following major categories:

1 Transient Neutropenia

Transient neutropenia is the most common form in neonates and typically resolves within days to weeks.

Causes include:

- **Maternal hypertension/preeclampsia:** Infants born to hypertensive mothers frequently develop neutropenia due to impaired placental function and reduced fetal neutrophil production (23).
- **Intrauterine growth restriction (IUGR):** Lineage steal phenomenon redirects hematopoiesis toward erythroid precursors at the expense of myeloid production (24).
- **Perinatal asphyxia and stress:** Severe perinatal stress can cause neutrophil margination and depletion of the circulating pool.
- **Acute bacterial or viral infections:** Sepsis, particularly with gram-negative organisms, and viral infections (including CMV, enteroviruses, and HIV) commonly suppress neutrophil counts (25).

2 Immune-Mediated Neutropenia

This category includes:

- **Alloimmune neonatal neutropenia (ANN):** Caused by maternal IgG antineutrophil antibodies crossing the placenta; analogous to hemolytic disease of the newborn (15).
- **Autoimmune neutropenia of infancy (AIN):** Driven by autoantibodies against neutrophil antigens; typically self-limited (18).
- **Maternal autoimmune neutropenia:** Infants of mothers with autoimmune neutropenia may acquire neutropenia via transplacental antibody transfer.

3 Congenital Bone Marrow Failure Syndromes

- **Severe congenital neutropenia (SCN):** Most commonly due to ELANE mutations; characterized by profound neutropenia from birth, maturation arrest at the promyelocyte/myelocyte stage, and recurrent severe infections (20).
- **Cyclic neutropenia (CyN):** Caused by specific ELANE mutations; characterized by regular 21-day oscillations of the ANC (26).
- **Shwachman-Diamond syndrome (SDS):** Caused by SBDS mutations; features neutropenia, exocrine pancreatic insufficiency, and skeletal abnormalities (27).
- **Other inherited marrow failure syndromes:** Including Fanconi anemia, dyskeratosis congenita, and cartilage-hair hypoplasia, which may present with neutropenia as part of a broader cytopenia profile.

3 Congenital Bone Marrow Failure Syndromes

- **Severe congenital neutropenia (SCN):** Most commonly due to ELANE mutations; characterized by profound neutropenia from birth, maturation arrest at the promyelocyte/myelocyte stage, and recurrent severe infections (20).
- **Cyclic neutropenia (CyN):** Caused by specific ELANE mutations; characterized by regular 21-day oscillations of the ANC (26).
- **Shwachman-Diamond syndrome (SDS):** Caused by SBDS mutations; features neutropenia, exocrine pancreatic insufficiency, and skeletal abnormalities (27).
- **Other inherited marrow failure syndromes:** Including Fanconi anemia, dyskeratosis congenita, and cartilage-hair hypoplasia, which may present with neutropenia as part of a broader cytopenia profile.

4 Drug-Induced Neutropenia

Several medications commonly used in neonates can cause neutropenia, including:

- **Antibiotics:** Beta-lactams (particularly piperacillin-tazobactam), vancomycin, linezolid, and chloramphenicol (28).
- **Antifungals:** Flucytosine and high-dose amphotericin B.
- **Antivirals:** Ganciclovir and valganciclovir, particularly in the treatment of congenital CMV infection.
- **Other drugs:** Ibuprofen, indomethacin, and ranitidine have been implicated.

5 Nutritional and Metabolic Causes

- **Folate deficiency:** Particularly in infants fed goat's milk or those with malabsorption.
- **Vitamin B12 deficiency:** In infants of vegan mothers or those with pernicious anemia.
- **Copper deficiency:** Associated with neutropenia and is seen in malnourished infants or those receiving prolonged parenteral nutrition without adequate copper supplementation (29).

6 Severity-Based Classification

The European Hematology Association (EHA) and EuNet-INNOCHRON consensus guidelines classify neutropenia severity based on ANC values (1):

- Mild neutropenia: ANC $1.0\text{--}1.5 \times 10^9/\text{L}$
- Moderate neutropenia: ANC $0.5\text{--}0.99 \times 10^9/\text{L}$
- Severe neutropenia: ANC $0.2\text{--}0.49 \times 10^9/\text{L}$
- Very severe neutropenia: ANC $< 0.2 \times 10^9/\text{L}$

This severity grading informs the urgency of intervention, with very severe neutropenia requiring immediate hematology consultation and consideration of empirical antimicrobial therapy.

Pathophysiology

The pathophysiology of infant neutropenia can be conceptualized along three principal mechanisms: impaired production, increased destruction, and altered distribution or sequestration.

1 Impaired Production

Impaired neutrophil production is the hallmark of congenital bone marrow failure syndromes. In ELANE-related SCN, mutations in the ELANE gene (encoding neutrophil elastase) cause misfolding of the enzyme, triggering endoplasmic reticulum stress and accelerated apoptosis of myeloid progenitor cells (30). The bone marrow characteristically shows maturation arrest at the promyelocyte or myelocyte stage, with reduced or absent mature neutrophils.

Similarly, HAX1 mutations (the genetic basis of classic Kostmann disease) disrupt mitochondrial homeostasis, leading to enhanced apoptosis of myeloid precursors (31). G6PC3 deficiency impairs glucose-6-phosphatase function, causing accumulation of toxic metabolites and neutrophil apoptosis (32).

In preterm infants, impaired production reflects the developmental immaturity of the myeloid progenitor pool. The bone marrow of preterm neonates contains a smaller proportion of committed myeloid progenitors compared with term neonates, and the storage pool of mature neutrophils is reduced, making these infants particularly vulnerable to neutropenia during periods of increased demand (33).

2 Increased Destruction

Immune-mediated neutropenia results from antibody-mediated destruction of circulating neutrophils. In alloimmune neonatal neutropenia, maternal IgG antibodies directed against fetal neutrophil antigens (most commonly HNA-1a) opsonize fetal neutrophils, leading to their destruction by the reticuloendothelial system (15). In autoimmune neutropenia of infancy, autoantibodies, typically IgG, coat the infant's own neutrophils, with similar downstream consequences. The bone marrow in these conditions typically shows normal or increased myeloid cellularity with appropriate maturation, reflecting a compensatory response to peripheral destruction.

3 Altered Distribution and Sequestration

In certain conditions, neutrophils marginate from the circulating pool into the marginal pool (adhering to vascular endothelium), leading to a reduced ANC without true depletion of total body neutrophils. This mechanism contributes to neutropenia in acute infections, stress states, and after administration of certain drugs (34). The distinction between true neutropenia and pseudoneutropenia due to margination can be challenging but is important for clinical decision-making.

4 Molecular Pathways and Emerging Insights

Recent research has elucidated additional molecular pathways involved in neutrophil homeostasis. The CSF3R gene encodes the G-CSF receptor, and mutations causing receptor truncation lead to impaired response to endogenous G-CSF and severe neutropenia (35). GFI1 acts as a transcriptional repressor that modulates granulopoiesis, with mutations causing both neutropenia and monocytosis (36). The WAS gene, classically associated with Wiskott-Aldrich syndrome, can also cause X-linked neutropenia without the full syndromic features (37).

The recognition that unfolded protein response (UPR) activation plays a central role in ELANE-related SCN has informed the development of novel therapeutic strategies. Gene therapy approaches targeting ELANE expression, using CRISPR-based base editing and prime editing, have demonstrated restoration of neutrophil maturation in preclinical models (11,12).

Clinical Manifestations

The clinical presentation of infant neutropenia is highly variable and depends primarily on the severity and duration of the neutropenia, as well as the underlying etiology.

1 Asymptomatic Presentation

Many infants with mild or moderate neutropenia are asymptomatic, with the abnormality detected incidentally on routine complete blood counts. This is particularly common in cases of transient neutropenia related to maternal hypertension, IUGR, or mild viral infections (38).

2 Infectious Complications

When neutropenia is severe ($ANC < 0.5 \times 10^9/L$) or prolonged, the risk of bacterial and fungal infections increases substantially. Common clinical presentations include:

- **Skin and soft tissue infections:** Cellulitis, abscesses, and impetigo, often with unusual organisms such as *Pseudomonas aeruginosa* or *Serratia marcescens* (39).
- **Omphalitis:** Infection of the umbilical stump, particularly in neonates.
- **Oral and mucosal infections:** Gingivitis, stomatitis, aphthous ulcers, and periodontitis, especially in cyclic neutropenia (26).
- **Pneumonia:** Bacterial pneumonia is a frequent and potentially severe complication.
- **Sepsis:** Gram-negative sepsis is a leading cause of mortality in severe neutropenia, particularly in neonates (40).
- **Fungal infections:** Candidemia and invasive aspergillosis may occur in prolonged severe neutropenia.

3 Characteristic Patterns by Etiology

- **Severe congenital neutropenia:** Presents in early infancy with recurrent, severe bacterial infections including deep-seated abscesses, pneumonia, and sepsis. Omphalitis and delayed umbilical cord separation (>30 days) are classic features (20).
- **Cyclic neutropenia:** Characterized by predictable 21-day cycles of fever, oral ulcers, gingivitis, and pharyngitis, with neutropenic nadirs lasting 3–5 days (26).
- **Autoimmune neutropenia of infancy:** Often presents with mild to moderate infections, including recurrent otitis media, upper respiratory infections, and skin infections. Severe infections are uncommon (18).
- **Alloimmune neonatal neutropenia:** May be asymptomatic or present with mild infections; severe infections occur in approximately 10–20% of cases (15).
- **Shwachman-Diamond syndrome:** Presents with failure to thrive due to pancreatic insufficiency, recurrent infections from neutropenia, and skeletal abnormalities (27).

4 Associated Features Suggesting Specific Diagnoses

- **Delayed umbilical cord separation:** Suggests leukocyte adhesion deficiency but can also occur in severe congenital neutropenia.
- **Congenital anomalies (cardiac, urogenital, vascular):** Suggest G6PC3-related SCN (32).
- **Exocrine pancreatic insufficiency and skeletal dysplasia:** Suggest Shwachman-Diamond syndrome (27).
- **Skin pigmentation anomalies, thumb abnormalities:** Suggest Fanconi anemia.
- **Oral ulcers with cyclic pattern:** Highly suggestive of cyclic neutropenia (26).

Diagnostic approach

The diagnostic evaluation of an infant with neutropenia requires a systematic, stepwise approach that integrates clinical assessment, laboratory testing, and, when indicated, genetic and bone marrow evaluation.

1 Initial Clinical Assessment

A thorough history should include:

- **Perinatal history:** Gestational age, birth weight, maternal health (hypertension, preeclampsia, autoimmune disease), and any maternal medications.
- **Family history:** Consanguinity, recurrent infections in siblings, early childhood deaths, or known inherited marrow failure syndromes.
- **Infection history:** Frequency, severity, and types of infections; pattern of recurrence.
- **Medication exposure:** Recent or current medications, including those administered to the mother during pregnancy or lactation.
- **Nutritional history:** Feeding type, vitamin supplementation, and any signs of malabsorption.

Physical examination should assess for:

- Growth parameters and signs of failure to thrive
- Dysmorphic features or congenital anomalies
- Hepatomegaly, splenomegaly, or lymphadenopathy
- Active infections or their sequelae
- Features of syndromic bone marrow failure (skin pigmentation, nail abnormalities, skeletal anomalies)

2 Laboratory Evaluation

Complete Blood Count (CBC) with Differential Serial CBCs (at least 2–3 measurements over 1–2 weeks) are essential to confirm persistent neutropenia and to assess for cyclic patterns. The differential should include not only the ANC but also the absolute counts of lymphocytes, monocytes, eosinophils, and basophils, as abnormalities in other lineages may suggest broader marrow failure (41).

Peripheral Blood Smear

Morphologic examination can reveal:

- Left-shifted myeloid series suggesting infection
- Giant granules (Chédiak-Higashi syndrome)
- Hypogranular neutrophils (specific granule deficiency)
- Dysplastic features suggesting MDS

Antineutrophil Antibody Testing

Granulocyte agglutination test (GAT) and granulocyte immunofluorescence test (GIFT) can detect antineutrophil antibodies, supporting the diagnosis of immune-mediated neutropenia.

However, these tests have limited sensitivity (approximately 50–60%) and specificity, and a negative result does not exclude immune-mediated neutropenia (42).

Infection Evaluation

Blood cultures, urine cultures, and viral studies (CMV PCR, parvovirus B19, enterovirus) should be obtained based on clinical suspicion. Inflammatory markers (CRP, procalcitonin) may aid in identifying occult infection.

Immunoglobulin Levels

Quantitative immunoglobulins can help exclude combined immunodeficiency syndromes that may present with neutropenia.

3 Bone Marrow Examination

Bone marrow aspiration and biopsy are indicated when:

- Neutropenia is severe ($ANC < 0.5 \times 10^9/L$) and persistent (>2–4 weeks)
- There are additional cytopenias suggesting marrow failure
- Clinical features suggest inherited bone marrow failure syndrome
- Initial workup is inconclusive and the patient is symptomatic
- There is concern for malignant transformation (43)

The bone marrow in SCN characteristically shows maturation arrest at the promyelocyte/myelocyte stage, with reduced or absent mature neutrophils and bands. In immune-mediated neutropenia, the marrow typically shows normal or increased cellularity with appropriate maturation. Cytogenetic analysis and next-generation sequencing (NGS) of the marrow can detect acquired somatic mutations associated with MDS/AML (44).

4 Genetic Testing

Next-generation sequencing (NGS) panels targeting known neutropenia-associated genes (ELANE, HAX1, G6PC3, CSF3R, WAS, GFI1, SBDS, SRP54, DNAJC21, EFL1, VPS45, JAGN1, AK2, and others) have unexplained neutropenia (45). The European Hematology Association guidelines recommend genetic testing for all patients with chronic neutropenia of unknown etiology after initial workup (1).

Whole-exome sequencing (WES) or whole-genome sequencing (WGS) may be considered when targeted panels are non-diagnostic, particularly in syndromic presentations.

5 Diagnostic Algorithm

A practical diagnostic algorithm, adapted from the AIEOP (Italian Association of Pediatric Hematology-Oncology) and EHA guidelines, proceeds as follows:

1. Confirm persistent neutropenia with serial CBCs
2. Review perinatal and family history; perform physical examination
3. Obtain peripheral blood smear and antineutrophil antibody testing
4. Evaluate for infections and drug-induced causes
5. If persistent, severe, or unexplained: proceed to bone marrow examination and genetic testing
6. If genetic testing is non-diagnostic: consider WES/ WGS

Differential Diagnosis

The differential diagnosis of infant neutropenia is broad and requires consideration of multiple categories. A structured approach based on duration, severity, and associated features can help narrow the differential.

1 Transient versus Persistent Neutropenia

Transient neutropenia is suggested by:

- Recent perinatal complication (hypertension, asphyxia, IUGR)
- Concurrent infection
- Recent medication exposure
- Resolution within days to weeks

Persistent neutropenia (≥ 4 weeks) warrants evaluation for:

- Immune-mediated causes (AIN, ANN)
- Congenital bone marrow failure syndromes
- Chronic infections (HIV, CMV)
- Nutritional deficiencies

2 Isolated versus Syndromic Neutropenia

Isolated neutropenia (with normal other cell lines) suggests:

- Immune-mediated neutropenia
- ELANE-related SCN
- Cyclic neutropenia
- Drug-induced neutropenia

Syndromic neutropenia (with other cytopenias or congenital anomalies) suggests:

- Shwachman-Diamond syndrome (pancytopenia, pancreatic insufficiency, skeletal anomalies)
- Fanconi anemia (pancytopenia, thumb anomalies, skin pigmentation)
- Dyskeratosis congenita (pancytopenia, nail dystrophy, mucosal leukoplakia)
- G6PC3-related SCN (cardiac and urogenital anomalies)

3 Severe versus Mild Neutropenia

Severe neutropenia ($ANC < 0.5 \times 10^9/L$) is more commonly associated with:

- Severe congenital neutropenia
- Alloimmune neonatal neutropenia (severe cases)
- Acute severe infections

Mild to moderate neutropenia ($ANC 0.5\text{--}1.5 \times 10^9/L$) is more commonly associated with:

- Autoimmune neutropenia of infancy
- Transient physiologic neutropenia
- Ethnic neutropenia (benign ethnic neutropenia)

4 Cyclic Pattern

A cyclic pattern with regular 21-day oscillations strongly suggests cyclic neutropenia due to ELANE mutations. Serial CBCs (2–3 times weekly for 4–6 weeks) can establish the cyclic pattern before genetic confirmation (26).

Management and Treatment

Management of infant neutropenia is tailored to the underlying etiology, severity, and presence of symptoms. The overarching goals are to prevent and treat infections, restore safe neutrophil counts when possible, and address the underlying cause.

1 General Principles

- **Infection prevention:** Hand hygiene, avoidance of sick contacts, and prompt evaluation of fever.
- **Prophylactic antibiotics:** The role of routine prophylactic antibiotics (e.g., trimethoprim-sulfamethoxazole) is controversial; some centers use them in severe neutropenia, while others reserve them for patients with recurrent infections despite G-CSF therapy (46).
- **Antifungal prophylaxis:** Considered in patients with prolonged severe neutropenia or those receiving high-dose G-CSF.
- **Vaccination:** Routine immunizations are generally safe; live vaccines should be avoided during severe neutropenia or active infection.

2 Management of Transient Neutropenia

Transient neutropenia typically requires no specific therapy beyond monitoring and supportive care. If the infant is well and ANC is $>1.0 \times 10^9/L$, no intervention is needed. For moderate neutropenia ($0.5\text{--}0.99 \times 10^9/L$), serial monitoring and parental education about signs of infection are appropriate. Severe neutropenia ($<0.5 \times 10^9/L$) in a symptomatic infant warrants hospitalization, empirical antibiotics, and hematology consultation (47).

3 Management of Immune-Mediated Neutropenia

Alloimmune Neonatal Neutropenia

- Most cases are self-limited and resolve within weeks as maternal antibodies wane.
- Empirical antibiotics for suspected bacterial infections.
- G-CSF ($5\text{--}10 \mu\text{g/kg/day}$ subcutaneously) is effective in accelerating neutrophil recovery in severe cases (48).
- Recurrence risk in subsequent pregnancies is approximately 25%; antenatal management may be considered in severe cases.

Autoimmune Neutropenia of Infancy

- Most cases are mild and require only observation.
- Antibiotics for documented infections.
- G-CSF reserved for severe or recurrent infections; typically produces rapid ANC response.
- The condition usually resolves spontaneously within 2–4 years (18).

4 Management of Severe Congenital Neutropenia

Granulocyte Colony-Stimulating Factor (G-CSF)

G-CSF is the cornerstone of therapy for SCN. The landmark studies by Bonilla et al. and the Severe Chronic Neutropenia International Registry (SCNIR) demonstrated that daily subcutaneous G-CSF (typically starting at $5 \mu\text{g/kg/day}$ and titrated to maintain ANC $>1.0 \times 10^9/L$) dramatically reduces infection frequency and improves survival (10,49). Approximately 90% of patients respond to G-CSF, with most achieving target ANC values at doses of $5\text{--}10 \mu\text{g/kg/day}$. Higher doses ($>20 \mu\text{g/kg/day}$) are associated with increased risk of MDS/AML and may prompt consideration of hematopoietic stem cell transplantation (HSCT) (50).

Hematopoietic Stem Cell Transplantation (HSCT)

HSCT is the only curative therapy for SCN and is indicated for:

- Patients with poor response to G-CSF (requiring $>20 \mu\text{g/kg/day}$)
- Patients with evidence of clonal evolution or MDS/AML
- Patients with severe infections despite optimal G-CSF therapy (51)

Outcomes of HSCT in SCN have improved substantially in recent years, with overall survival rates exceeding 80% in matched sibling donor transplants (52).

Infection Management

- Prompt empirical antibiotics for fever
- Consideration of antifungal therapy for prolonged fever
- Aggressive evaluation and treatment of focal infections

5 Management of Cyclic Neutropenia

G-CSF is highly effective in cyclic neutropenia, reducing the depth and duration of neutropenic nadirs and preventing symptomatic cycles. Low-dose G-CSF ($1\text{--}3 \mu\text{g/kg/day}$) is often sufficient, and most patients achieve excellent control with minimal side effects (26).

6 Management of Shwachman-Diamond Syndrome

Management of SDS is multifaceted and includes:

- Pancreatic enzyme replacement therapy for exocrine insufficiency
- Fat-soluble vitamin supplementation
- G-CSF for severe neutropenia (with caution due to MDS/AML risk)
- HSCT for progressive marrow failure or leukemic transformation (27)

7 Management of Drug-Induced Neutropenia

- Discontinuation of the offending agent
- Supportive care with monitoring of ANC recovery
- G-CSF in severe or prolonged cases

8 Management of Nutritional Neutropenia

- Folate or vitamin B12 supplementation as appropriate
- Copper supplementation for copper deficiency
- Resolution typically occurs within days to weeks of supplementation

9 Role of Biosimilars

Biosimilar G-CSF products (e.g., filgrastim-sndz) have demonstrated comparable efficacy and safety to reference filgrastim in pediatric populations, including for post-HSCT engraftment and chemotherapy-induced neutropenia prophylaxis (53). The availability of biosimilars has reduced treatment costs and improved access to G-CSF therapy.

Prognosis

The prognosis of infant neutropenia is highly dependent on the underlying etiology.

1 Transient Neutropenia

Prognosis is excellent, with complete resolution typically occurring within days to weeks.

No long-term sequelae are expected.

2 Autoimmune Neutropenia of Infancy

Prognosis is favourable, with spontaneous resolution in approximately 90% of cases within 2–4 years (18).

Long-term follow-up studies have demonstrated low risk of severe late infections after resolution. Some patients may develop other autoimmune conditions, although this association is not firmly established.

3 Alloimmune Neonatal Neutropenia

Prognosis is generally good, with resolution within weeks as maternal antibodies are cleared. However, severe infections (particularly gram-negative sepsis) can be fatal in a minority of cases (15). Recurrence in subsequent pregnancies is a consideration for family counseling.

4 Severe Congenital Neutropenia

Prognosis has improved dramatically with G-CSF therapy. Pre-G-CSF era mortality was approximately 50% within the first year of life; modern series report survival rates exceeding 90% with G-CSF therapy (49). However, the risk of malignant transformation remains a major concern:

- **MDS/AML risk:** Approximately 20–25% of patients with SCN develop MDS or AML, with risk correlated with G-CSF dose and duration (50).
- **High-risk mutations:** Specific ELANE mutations (e.g., G214R, L121H, L121P) and CSF3R truncation mutations confer higher leukemia risk (54).
- **Monitoring:** Regular bone marrow examination (every 1–3 years) with cytogenetic analysis and NGS is recommended to detect early clonal evolution (44).

5 Cyclic Neutropenia

Prognosis is excellent with G-CSF therapy. Unlike SCN, cyclic neutropenia carries minimal risk of malignant transformation, with only rare cases of AML reported (54).

6 Shwachman-Diamond Syndrome

Prognosis is variable and depends on the severity of marrow failure and the risk of leukemic transformation. Approximately 10–20% of patients develop MDS or AML, with the highest risk in those with biallelic SBDS mutations and complex cytogenetic abnormalities (27).

Future Perspectives

The landscape of neutropenia management is evolving rapidly, with several promising developments on the horizon.

1 Gene Therapy and Gene Editing

CRISPR-based gene editing technologies have emerged as potential curative therapies for monogenic neutropenia syndromes. Recent preclinical studies have demonstrated:

- **Base editing and prime editing of ELANE:** Near-universal correction of ELANE mutations in CD34+ hematopoietic stem/progenitor cells, with restoration of neutrophil maturation and function in murine models (11).
- **Adenine base editing to inhibit mutant ELANE expression:** Selective silencing of the mutant allele while preserving wild-type function, offering a potentially safer alternative to nuclease-mediated editing (12).
- **Patient-cell-based gene therapy approaches:** Preclinical models demonstrating long-term disease correction (55).

These approaches hold promise for translation into clinical trials within the next several years, potentially offering curative treatment without the risks of allogeneic HSCT.

2 Improved Diagnostics

- **Expanded NGS panels:** Including newly identified neutropenia-associated genes (e.g., VPS45, JAGN1, AK2, SRP54).
- **Whole-genome sequencing:** Increasingly accessible and may identify novel genetic causes in currently unexplained cases.
- **Functional assays:** Including in vitro neutrophil differentiation assays and UPR activation studies to characterize variant pathogenicity.

3 Refined Risk Stratification

Ongoing research aims to refine leukemia risk prediction in SCN through:

- Integration of genetic data (specific ELANE mutations, CSF3R status)
- Somatic mutation monitoring (e.g., CSF3R, RUNX1, ASXL1)
- Machine learning approaches incorporating clinical and molecular data

4 Novel Therapeutics

- **Small molecule approaches:** Targeting endoplasmic reticulum stress in ELANE-related SCN.
- **Anti-apoptotic agents:** Investigated in preclinical models of HAX1-related neutropenia.
- **Improved G-CSF formulations:** Including longer-acting pegylated forms (pegfilgrastim) and biosimilars.

5 International Collaboration and Registries

International registries, including the Severe Chronic Neutropenia International Registry (SCNIR) and the European Net-INNOCHRON network, continue to play a critical role in advancing understanding of neutropenia through standardized data collection, genotype-phenotype correlation studies, and clinical trial infrastructure (1,49).

Conclusion

Neutropenia in infants encompasses a heterogeneous spectrum of disorders ranging from transient, self-limited conditions to severe, life-threatening congenital syndromes. Accurate diagnosis requires a systematic approach integrating clinical assessment, serial hematologic evaluation, antineutrophil antibody testing, bone marrow examination, and increasingly, comprehensive genetic testing. The advent of molecular diagnostics has transformed the classification and prognostication of inherited neutropenia syndromes, enabling personalized management strategies.

Recombinant human G-CSF remains the cornerstone of therapy for severe and symptomatic neutropenia, dramatically reducing infection-related morbidity and mortality. The recognition of long-term risks, particularly the association between high-dose G-CSF therapy and malignant transformation in SCN, has prompted refined treatment algorithms and careful monitoring protocols. Hematopoietic stem cell transplantation offers curative potential for patients with poor G-CSF response or evidence of clonal evolution.

Emerging gene editing technologies, particularly CRISPR-based base and prime editing approaches targeting ELANE and other neutropenia-associated genes, hold promise for definitive treatment of monogenic neutropenia syndromes. Continued international collaboration through registries and clinical networks will be essential to translate these advances into improved outcomes for affected infants.

Clinicians caring for infants with neutropenia must balance the imperative to prevent and treat infections with the need to avoid unnecessary interventions in self-limited conditions. A nuanced, evidence-based approach, guided by contemporary international guidelines and emerging molecular insights, offers the best path to optimizing outcomes in this vulnerable population.

References

1. Fioredda F, Skokowa J, Tamary H, et al. The European Guidelines on Diagnosis and Management of Neutropenia in Adults and Children: A Consensus Between the European Hematology Association and the EuNet-INNOCHRON COST Action. *Hemasphere*. 2023;7(4):e872.
2. Donadieu J, Fenneteau O, Beaupain B, et al. Congenital neutropenia: diagnosis, molecular bases and patient management. *Orphanet J Rare Dis*. 2011;6:26.
3. Schmutz N, Henry E, Jopling J, Christensen RD. Expected ranges for blood neutrophil concentrations of neonates: the Manroe and Mouzinho charts revisited. *J Perinatol*. 2008;28(4):275-281.
4. Christensen RD, Henry E, Jopling J, Wiedmeier SE. The CBC: reference ranges for neonates. *Semin Perinatol*. 2009;33(1):3-11.
5. Maheshwari A. Neutropenia in the newborn. *Indian J Pediatr*. 2008;75(4):371-374.
6. Boxer LA. Severe congenital neutropenia: genetics and pathogenesis. *Trans Am Clin Climatol Assoc*. 2006;117:47-63.
7. Rosenberg PS, Alter BP, Bolyard AA, et al. The incidence of leukemia and mortality from sepsis in patients with severe congenital neutropenia receiving long-term G-CSF therapy. *Blood*. 2006;107(1):46-51.
8. Skokowa J, Dale DC, Touw IP, et al. Severe congenital neutropenias. *Nat Rev Dis Primers*. 2017;3:17032.
9. Boztug K, Klein C. Genetics and pathophysiology of severe congenital neutropenia syndromes. *Curr Opin Hematol*. 2020;27(1):32-39.
10. Bonilla MA, Gillio AP, Ruggeiro M, et al. Effects of recombinant human granulocyte colony-stimulating factor on neutropenia in patients with congenital agranulocytosis. *N Engl J Med*. 1989;320(24):1574-1580.
11. Efficient, potent, and specific near-universal base and prime editing for ELANE-mutant severe congenital neutropenia. *Blood*. 2025;144(Suppl 1):624.
12. Targeted inhibition of ELANE expression using adenine base editing to treat severe congenital neutropenia. *Mol Ther Methods Clin Dev*. 2025;36:101626.
13. Funke A, Berner R, Traichel B, et al. Frequency, natural course, and outcome of neonatal neutropenia. *Pediatrics*. 2000;106(4):E53.
14. Carr R, Modi N, Dore CJ. G-CSF and GM-CSF for treating or preventing neonatal sepsis. *Cochrane Database Syst Rev*. 2003;(3):CD003066.
15. Curtis BR, Reno C, Aster RH. Neonatal alloimmune neutropenia attributed to maternal alloimmunization against human neutrophil alloantigen-1 and -3. *Blood*. 2014;124(21):4948.
16. van den Tooren-de Groot R, Ottink MD, Huisman EJ, et al. Management and outcome of neonatal alloimmune neutropenia. *Vox Sang*. 2022;117(10):1148-1156.
17. de Haas M, Muniz-Diaz E, Gaspar HB, et al. Neonatal alloimmune neutropenia: a Brazilian cross-sectional study. *Blood*. 2014;124(21):4948.
18. Farruggia P, Fioredda F, Puccio G, et al. Autoimmune neutropenia of infancy: clinical course and long-term outcome. *Br J Haematol*. 2019;186(4):e80-e83.

19. Dale DC, Person RE, Bolyard AA, et al. Mutations in the gene encoding neutrophil elastase (ELA2) are associated with severe congenital neutropenia and cyclic neutropenia. *Nat Genet.* 2000;24(4):413-417.
20. Makaryan V, Zeidler C, Bolyard AA, et al. Novel case of ELANE-related severe congenital neutropenia with initially normal bone marrow. *Pediatr Blood Cancer.* 2023;70(8):e30468.
21. Hsieh MM, Everhart JE, Byrd-Holt DD, et al. Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. *Ann Intern Med.* 2007;146(7):486-492.
22. Droma Y, McCullough RG, McCullough RE, et al. Changes in neutrophil and lymphocyte counts in high-altitude residents. *Wilderness Environ Med.* 2003;14(4):225-229.
23. Koenig JM, Christensen RD. Incidence, neutrophil kinetics, and natural history of neonatal neutropenia associated with maternal hypertension. *Pediatrics.* 1989;84(3):432-437.
24. Christensen RD, Yoder BA, Baer VL, et al. Early-onset neutropenia in small-for-gestational-age infants. *Pediatrics.* 2006;118(5):e1503-e1510.
25. Al-Hadidi SA, Abutaleb A, Aljadani M, et al. Neonatal sepsis and neutropenia: a review. *J Matern Fetal Neonatal Med.* 2022;35(25):9020-9027.
26. Makaryan V, Boulad F, Epling-Burnette PK, et al. Cyclic neutropenia: a review. *Orphanet J Rare Dis.* 2020;15(1):328.
27. Myers KC, Furutani E, Tsang J, et al. Clinical characteristics and genetic mutation analysis in 18 pediatric patients with Shwachman-Diamond syndrome. *Front Genet.* 2025;16:1603782.
28. Andersohn F, Konzen C, Garbe E. Systematic review: agranulocytosis induced by nonchemotherapy drugs. *Ann Intern Med.* 2007;146(9):657-665.
29. Cordano A, Graham GG. Copper deficiency complicating severe chronic intestinal malabsorption. *Pediatrics.* 1966;38(4):596-604.
30. Horwitz MS, Duan Z, Korkmaz B, et al. Neutrophil elastase in cyclic and severe congenital neutropenia. *Blood.* 2007;109(5):1817-1824.
31. Klein C, Grudzien M, Appaswamy G, et al. HAX1 deficiency causes autosomal recessive severe congenital neutropenia (Kostmann disease). *Nat Genet.* 2007;39(1):86-92.
32. Boztug K, Appaswamy G, Ashikov A, et al. A syndrome with congenital neutropenia and mutations in G6PC3. *N Engl J Med.* 2009;360(1):32-43.
33. Carr R, Huizinga TW. Low-signal neutrophils in preterm neonates. *J Pediatr.* 2000;137(4):560-565.
34. Athens JW, Haab OP, Raab SO, et al. Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *J Clin Invest.* 1961;40(6):989-995.
35. Germeshausen M, Ballmaier M, Welte K. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: results of a long-term survey of 96 patients. *Blood.* 2006;107(1):46-51.
36. Person RE, Li FQ, Duan Z, et al. Mutations in proto-oncogene GFI1 cause human neutropenia. *Nat Genet.* 2003;34(3):308-312.
37. Devriendt K, Kim AS, Mathijs G, et al. Constitutively activating mutation in WASP causes X-linked severe congenital neutropenia. *Nat Genet.* 2001;27(3):313-317.
38. Watts RG. Neutropenia. In: *Nelson Textbook of Pediatrics.* 22nd ed. Philadelphia: Elsevier; 2024:chap 159.
39. Bodey GP, Buckley M, Sathe YS, Freireich EJ. Quantitative relationships between circulating leukocytes and infection in patients with acute leukemia. *Ann Intern Med.* 1966;64(2):328-340.
40. Klastersky J, Ameye L, Maertens J, et al. Bacteraemia in febrile neutropenic cancer patients. *Int J Antimicrob Agents.* 2007;30(Suppl 1):S51-S59.
41. Dale DC. How I manage children with neutropenia. *Br J Haematol.* 2017;178(3):351-363.
42. Cattaneo M, Biffi A, Notarangelo LD, et al. Laboratory testing for the diagnosis of neonatal and pediatric immune neutropenias: a narrative review. *Pediatr Med.* 2020;3:14.
43. Fioredda F, Calvillo M, Bonanomi S, et al. Diagnosis and management of neutropenia in children: The approach of the Study Group on Neutropenia and Marrow Failure Syndromes of the Pediatric Italian Hemato-Oncology Association (AIEOP). *Pediatr Blood Cancer.* 2022;69(4):e29599.
44. Beekman JM, Schaik RH, van Reijssen M, et al. Monitoring MDS/AML evolution in severe congenital neutropenia patients treated with G-CSF. *Blood.* 2020;136(Suppl 1):23-24.
45. Chou JY, Kaufman RE. Severe congenital neutropenia: a single-center experience. *J Pediatr Hematol Oncol.* 2021;43(8):e1142-e1147.
46. van de Vijver E, van den Berg TK, Kuijpers TW. Prophylactic antibiotics in neutropenic patients. *N Engl J Med.* 2013;369(15):1472-1473.
47. NHS Lothian. Neonatal Neutropenia Guideline. NHS Scotland; 2021.
48. Maheshwari A, Christensen RD, Calhoun DA. Immune-mediated neutropenia in the neonate. *Acta Paediatr.* 2002;91(438):98-103.
49. Dale DC, Bolyard AA, Shannon JA, et al. Outcomes for patients with severe congenital neutropenia treated with granulocyte colony-stimulating factor. *Blood.* 2020;136(Suppl 1):33-34.
50. Rosenberg PS, Zeidler C, Bolyard AA, et al. Stable long-term risk of leukemia in patients with severe congenital neutropenia. *Blood.* 2020;136(15):1858-1864.
51. Connelly JA, Choi SW, Levine JE. Hematopoietic stem cell transplantation for severe congenital neutropenia. *Curr Opin Hematol.* 2020;27(1):40-46.
52. Fioredda F, Iacobelli S, van Biezen A, et al. Stem cell transplantation in severe congenital neutropenia: an analysis from the European Society for Blood and Bone Marrow Transplantation. *Blood.* 2015;126(17):2003-2007.
53. Park S, Kim HS, Choi JY, et al. Comparing the efficacy and safety of reference drug filgrastim and biosimilar filgrastim-sndz in pediatric patients. *J Hematol Oncol Pharm.* 2023;13(1):28-34.

54. Makaryan V, Zeidler C, Bolyard AA, et al. Clinical relevance of SCN and CyN induced by ELANE mutations: a systematic review. *Front Immunol.* 2024;15:1349919.
55. Seattle Children's Research Institute. Research shared at ASGCT 2025. Available at: <https://www.seattlechildrens.org/research/featured-research/asgct-2025/>. Accessed 2025.
56. ELANE-Related Neutropenia. GeneReviews® [Internet]. Seattle (WA): University of Washington; 2002-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1533/>
57. Shwachman-Diamond Syndrome. GeneReviews® [Internet]. Seattle (WA): University of Washington; 2008-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1756/>
58. Carr R, Brocklehurst P, Dore CJ, Modi N. Granulocyte colony-stimulating factor for preventing neonatal sepsis in neutropenic infants. *Cochrane Database Syst Rev.* 2019;(7):CD003066.
59. Zhang X, Wang Y, Li X, et al. A multicenter phase II trial of primary prophylactic PEG-rhG-CSF in pediatric patients with solid tumors and non-Hodgkin lymphoma. *Cancer Med.* 2024;13(16):e6079.
60. Kostmann R. Infantile genetic agranulocytosis (agranulocytosis infantilis hereditaria). *Acta Paediatr.* 1956;45(Suppl 105):1-78.

Invisible Curriculum, Unmeasured Outcomes: A Theory-Informed Analysis of Quality Assurance Gaps in Family Medicine Training

Hakan Yaman ⁽¹⁾, Aylin Yaman ⁽²⁾

(1) Prof. Dr. Hakan Yaman, Tez Medical; and American Hospital, Occupational Physician; Independent Scholar; Prof. Dr. (Emeritus), Akdeniz University, Faculty of Medicine; Antalya, Türkiye

2) Prof. Dr. Aylin Yaman, Istinye University, Faculty of Medicine, Dep. of Neurology, Istanbul

Correspondence:

Prof. Dr. Hakan YAMAN

Antalya, Türkiye

Email: hakanyam@yahoo.com

Received: August 2026. Accepted: September 2026; Published: October 2026.

Citation: Hakan Yaman, Aylin Yaman. Invisible Curriculum, Unmeasured Outcomes: A Theory-Informed Analysis of Quality Assurance Gaps in Family Medicine Training. . World Family Medicine. October 2026; 24(7): 66- 75

DOI: 10.5742/MEWFM.2026.241798

Abstract

Background: Family practice training is central to health system performance, yet its quality assurance mechanisms remain insufficiently aligned. Despite the widespread adoption of competency-based medical education (CBME), persistent gaps exist between defined competencies and their measurable outcomes. This study explores these gaps through a theory-informed analysis of quality assurance structures in family medicine training in Türkiye.

Methods: This study employed a theory-informed conceptual and interpretive design. A structured analytical framework was developed based on health policy analysis, CBME theory, and quality assurance models, including continuous quality improvement (CQI). Documentary sources, including international frameworks (WHO, WONCA), national curricula (TUKMOS), institutional reports, and accreditation documents were purposively selected. Analysis was conducted across four predefined domains: educational outcomes, assessment systems, accreditation structures, and feedback mechanisms, using a stepwise approach of mapping, structuring, cross-level comparison, and conceptual synthesis.

Results: The findings reveal a consistent pattern of structural misalignment across system levels. Competencies are clearly defined but insufficiently operationalised into measurable outcomes.

Assessment systems are diverse but lack standardisation and comparability in Türkiye. Accreditation processes are well established yet predominantly institution-focused rather than outcome-oriented. Feedback mechanisms are widely implemented but poorly integrated into continuous improvement cycles.

These disconnections culminate in a systemic “invisible curriculum–unmeasured outcomes gap,” whereby critical components of competence, such as contextual decision-making, longitudinal care, and professional identity formation remain insufficiently captured within formal evaluation systems.

Conclusion: Quality assurance in family practice training is conceptually robust but functionally fragmented. Addressing this challenge requires a shift from component-based approaches toward a system-integrated, feedback-driven model that aligns competencies, assessment, accreditation, and feedback across all levels, which in many systems are not instituted. Such a reconceptualisation is essential to ensure that competence is not only defined but also measurable, validated, and continuously improved in practice.

Keywords: Family practice training, Quality assurance, Competency-based medical education (CBME), Assessment systems, Accreditation, Feedback mechanisms, Hidden curriculum, Medical education, Primary care, Health systems

Introduction

Family practice training is fundamental to effective health systems, influencing physician competence, care accessibility, continuity, and equity (1–5). Strong primary care improves health outcomes, reduces costs, and enhances care coordination (6–8). Consequently, the quality of family practice training is a key determinant of health system performance (9,10).

Medical education has shifted from knowledge-based to competency-based, outcome-oriented models over the past two decades (11,7). In family practice, WONCA and the World Health Organization have defined competency frameworks covering clinical expertise, communication, professionalism, and community orientation (8,9). These frameworks prepare physicians for the complexity and continuity of primary care (5,10). Despite these advances, translating competency frameworks into standardised, measurable outcomes remains inconsistent (11,12). Assessment methods vary across institutions and often emphasise knowledge or simulated performance rather than real-world competence (13–15). Moreover, fragmented feedback systems limit continuous educational improvement (16,17).

This creates an “invisible curriculum—unmeasured outcomes gap,” whereby competency-based curricula lack coherent programme-level evaluation despite structured assessment systems (18–22). Digital health, including decision-support systems, telemedicine, and AI-enabled patients, further expands competency requirements to digital literacy and adaptive decision-making (23–27). However, quality assurance remains focused on accreditation and structural standards rather than educational outcomes (28–30). This study examines how quality in family practice training is conceptualised, operationalised, and governed (31).

State of the Art: Quality Assurance in Family Practice Training

Family medicine is recognised as a scientific discipline with its own knowledge base and educational framework (32). However, training remains heterogeneous across countries and institutions (33). Much competence develops through the “invisible curriculum,” including longitudinal patient care, experiential learning, clinical uncertainty, and professional identity formation (18,19,34–36). Contemporary education promotes continuous quality improvement through feedback-driven learning (16,37,38). However, family practice training and quality assurance remain rooted in pre-digital models, with emerging competencies inadequately integrated into curricula and assessment (4,39). Thus, the main challenge is not the absence of quality frameworks but their fragmented implementation, highlighting the need for integrated, feedback-driven quality assurance systems. (Table 1).

Domain	State of Knowledge	Structural Gap
Competencies	well defined	poorly measured
Assessment	diverse tools exist	lack of standardisation
Accreditation	well developed	institution-focused
Feedback	present	not system-integrated
Digital transformation	emerging	not embedded

Table 1. Synthesis of the State of the Art in Quality Assurance for Family Practice Training

Family practice training in Türkiye has expanded rapidly since the Health Transformation Program, supported by competency-based postgraduate curricula such as TUKMOS (2,40– 44). Nevertheless, training remains heterogeneous across institutions, with substantial variation in competency assessment and measurable outcomes despite nationally defined standards (12,40,41,43). Quality assurance continues to rely mainly on institutional accreditation rather than programme-level outcome evaluation (45).

Aim of the Study

The aim of this study was to critically examine how quality in family practice training is conceptualised, operationalised, and governed, and to identify the structural gaps that emerge between these domains.

Materials and Methods

Study Design

This study used a theory-informed conceptual and interpretive approach to examine quality assurance in family practice training, focusing on structural relationships and patterns of alignment and misalignment rather than evidence aggregation or quantification. The analysis was guided by an integrated framework combining health policy analysis (31),

Competency-Based Medical Education (CBME) (7,22), and quality assurance/continuous improvement models (ESG, CQI) (21,28). These frameworks informed four predefined analytical domains: educational outcomes, assessment systems, accreditation structures, and feedback mechanisms.

Each domain was operationalised through specific analytical questions (Table 2).

Domain	Analytical Focus	Key Questions
Educational outcomes	Definition and measurability of competencies	Are competencies clearly defined? Are they measurable?
Assessment systems	Tools and standardisation	How are competencies assessed? Are methods comparable?
Accreditation	Institutional vs program-level focus	Does accreditation evaluate learning outcomes or structures?
Feedback mechanisms	Integration and improvement cycles	Is feedback used for continuous quality improvement?

Table 2. Analytical Framework for Evaluating Quality Assurance in Family Practice Training

Purposefully selected documentary sources were analysed as analytical artefacts rather than datasets. Selection was based on relevance to family practice training, inclusion of competency or quality assurance frameworks, representation of policy, institutional, and programme perspectives, and public accessibility. Sources included national curricula, institutional accreditation reports, international frameworks (e.g., WHO, WONCA), and key medical education literature. All sources were fully documented in the reference list.

1. International Policy and Educational Frameworks

World Health Organization (WHO). Primary Health Care: Transforming Vision into Action (2018) (3)
 World Health Organization (WHO). Global Strategy on Digital Health 2020–2025 (2021) (25)
 WONCA Europe. The European Definition of General Practice/Family Medicine (2011) (8)
 WONCA. Global Standards for Postgraduate Family Medicine Education (46)
 ESG (ENQA). Standards and Guidelines for Quality Assurance in the European Higher Education Area (2015) (28)
 European Union. Directive 2005/36/EC on the Recognition of Professional Qualifications (47)

2. Türkiye National Policy and Training Documents

T.C. Sağlık Bakanlığı. Sağlıkta Dönüşüm Programı (2003 ve devam eden raporlar) (48)
 Tıpta Uzmanlık Kurulu (TUK). Aile Hekimliği Uzmanlık Eğitimi Çekirdek Müfredatı (TUKMOS) (43)
 Yükseköğretim Kalite Kurulu (YÖKAK). Kurumsal Akreditasyon Programı Rehberi (45)
 Türkiye Cumhuriyeti Sağlık Bakanlığı eğitim ve hizmet sunumuna ilişkin resmi dokümanlar (49)

3. Institutional Documents (Türkiye)

Akdeniz University Faculty of Medicine – Family Medicine Training Programme documents (50)
 İstanbul University – Postgraduate medical education documentation (51)
 Ege University – Programme structure and assessment reports (52)
 Ankara University – Residency training guidelines (53)
 Dokuz Eylül University – Educational programme documentation (54)
 Atatürk University – Training and evaluation framework (55)

(Where available, publicly accessible curriculum outlines, evaluation forms, and institutional quality reports were reviewed.)

4. Accreditation and Quality Assurance Reports

University Internal Evaluation Reports (KİDR) (56)
 YÖKAK Institutional Accreditation Reports (57)
 Institutional quality assurance documentation (programme-level where available)

5. Key Academic and Theoretical Sources

Walt & Gilson – Health policy analysis framework (31)
 CBME literature (Frank et al., Holmboe et al., Ten Cate) (7,11, 22)
 Assessment theory (van der Vleuten et al., Hodges) (13,14,20,21)
 Quality assurance and CQI (Batalden & Davidoff) (58)
 Hidden curriculum (Hafferty) (18)
 Professional identity formation (Cruess et al.) (36)
 Digital health and AI in medicine (Topol, WHO reports) (23,25)

Analytical Procedure

The analysis followed a structured four-step iterative process designed to ensure transparency, consistency, and analytical rigour (31,59,60,61):

- 1. Mapping:** Documents were reviewed to identify definitions of competencies, learning outcomes, quality indicators, assessment, accreditation, and feedback, mapping how quality was conceptualised across sources.
- 2. Structuring:** Extracted data were organised into four predefined domains: educational outcomes, assessment systems, accreditation structures, and feedback mechanisms. Similar concepts were grouped and reviewed iteratively.
- 3. Cross-level comparison:** Policy, institutional, and programme levels were compared to identify alignment and structural misalignment between competencies, assessment, accreditation and feedback.
- 4. Conceptual synthesis:** Findings were synthesised to identify recurring patterns, structural gaps, and invisible or unmeasured competencies, leading to a theory-informed model of quality assurance as a dynamic, feedback-driven, system-integrated process.

No statistical analysis was performed because this was a theory-informed conceptual study.

Instead, a structured, iterative interpretive approach was used to identify patterns and structural misalignments. Analytical rigor was ensured through a predefined framework (educational outcomes, assessment, accreditation, and feedback), iterative analysis, cross-level comparison, triangulation of policy, institutional, and academic sources, and transparent documentation of methods. The aim was to generate a theory-informed, system-level understanding rather than statistically generalisable findings.

Patient and Public Involvement (PPI statement)

No patients or the public were involved, as the study analysed publicly available documents.

Although dissemination of these findings will not include direct patient involvement, the findings aim to improve training systems and, ultimately, patient care quality.

Ethical Considerations

Ethical approval was not required for this study, as it involved the analysis of publicly accessible policy documents, institutional reports, and academic literature. No human participants, personal data, or confidential information were included.

Results

System-Level Patterns and Structural Gaps in Family Practice Training

The analysis of policy documents, training curricula, institutional reports, and international frameworks revealed a consistent pattern:

quality assurance in family practice training is conceptually well defined but structurally fragmented across system levels.

Rather than isolated deficiencies, the findings point to systemic misalignments between educational outcomes, assessment systems, accreditation processes, and feedback mechanisms.

1 Educational Outcomes: Defined but weakly Operationalised

All analysed sources describe family practice training within a competency-based framework. International standards (WONCA and the World Health Organization) define core competencies including clinical expertise, communication, professionalism, and community orientation (3,8,9). In Türkiye, the TUKMOS curriculum closely aligns with these standards (7,43). Likewise, training programmes at major universities demonstrate formal consistency with national and international competency frameworks.

However, despite this high degree of conceptual alignment, a recurring pattern emerges across all sources:

competencies are clearly defined, but their operationalisation into measurable and standardised outcomes remains limited

In most documents:

learning objectives are described in broad and descriptive terms explicit performance indicators are rarely specified measurable benchmarks for competence are inconsistently defined.

Although competency frameworks are well established, their translation into measurable, standardised, and reproducible outcomes remains limited (11,22). Across all sources, competencies are clearly defined, but performance indicators and assessment criteria are inconsistently operationalised, creating a gap between competency definition and verification. This finding is consistent with the broader CBME literature on challenges in achieving valid and comparable assessment outcomes (11,20–22).

2 Assessment Systems: Diverse but Non-Comparable

All analysed programmes employ multiple assessment methods, including OSCEs, portfolios, and workplace-based tools such as the Mini-CEX, reflecting a shift toward multidimensional competency assessment (7,14,15). These approaches evaluate knowledge, clinical skills, and professional behaviour while improving alignment between assessment and real-world primary care practice (14,21).

However, despite this apparent methodological richness, a consistent pattern emerges across the analysed programmes:

assessment systems are diverse in form but limited in comparability and integration

Substantial variability exists in:

- the design and structure of assessment tools
- the frequency and timing of evaluations
- the criteria used to judge performance
- the documentation and use of assessment outcomes

Assessment practices vary across institutions despite using similar tools such as OSCEs and Mini-CEX, limiting comparability. Their validity is further constrained by assessor variability, contextual influences, and non-standardised scoring (14,20). Likewise, OSCEs and portfolios assess selected competencies but do not consistently reflect real-world clinical performance (15,34).

Taken together, these findings suggest that current assessment systems:

measure multiple aspects of performance but do not function as **standardised, comparable, or system-level indicators of competence**

This reinforces a broader limitation identified in the CBME literature: assessment systems, although conceptually aligned with competency frameworks, often lack the coherence and integration required to ensure robust and reproducible evaluation of competence across training settings (7,20,21).

Cross-document analysis showed marked variability in assessment design, frequency, performance criteria, and documentation across programmes. Although assessment systems exist universally, they lack standardisation and programmatic integration, limiting comparability and their contribution to quality improvement. Consequently,

assessments often measure activities rather than validated competence, consistent with challenges reported in CBME literature (14,20,21,34).

3 Accreditation Systems: Institution-Focused, Not Outcome-Focused

Accreditation analysis showed that quality assurance in family practice training is primarily institution-based, focusing on governance, strategic planning, documentation, and compliance with predefined standards (28,29). These mechanisms support institutional accountability and consistency.

However, a key limitation emerges across all analysed sources:

accreditation evaluates structures rather than educational outcomes

Across multiple institutional reports from major universities in Türkiye (including Akdeniz, Hacettepe, Ege, İstanbul, Ankara, and Gazi Universities), several consistent patterns are observed:

programme-level learning outcomes are **not systematically measured** the **effectiveness of assessment systems** is rarely evaluated

explicit links between **training processes and clinical competence** are not clearly demonstrated.

Accreditation provides limited evidence that training programmes produce competent practitioners, reflecting a broader gap between institutional compliance and educational effectiveness (29). In family practice, where competence is context-dependent and longitudinal, structural indicators alone are insufficient. Thus, accreditation ensures minimum standards but does not validate real-world training outcomes.

4 Feedback Mechanisms: Present but Not System-Integrated

All analysed institutions incorporate feedback mechanisms, including student surveys, course evaluations, and institutional review reports, recognising feedback as a formal component of quality assurance and educational monitoring.

However, the analysis reveals a consistent limitation:

Feedback is predominantly collected, but not systematically integrated into educational improvement processes.

Feedback is predominantly perception-based, emphasising learner satisfaction over objective clinical performance. Competence-based feedback is limited and inconsistently integrated into institutional evaluation and decision-making processes.

Most importantly, feedback systems are not embedded within **closed-loop quality improvement cycles**. There is limited evidence that: feedback data is systematically used to inform curriculum redesign performance data drives institutional or program-level change improvements are tracked longitudinally over time.

As a result, feedback mechanisms tend to function as stand-alone data collection tools rather than as drivers of continuous quality improvement. In this context, feedback remains: **collected but not operationalised**.

This pattern aligns with the broader literature, which emphasises that the effectiveness of feedback depends not on its presence, but on its integration into adaptive, system-level learning processes (16,17,21).

5 Cross-Level Misalignment: The Core Structural Problem

Cross-level analysis revealed systematic misalignment: competencies are defined at the policy level, assessments at the programme level, accreditation at the institutional level, and feedback at the learner level, but these components remain poorly integrated. Competency frameworks are not consistently translated into assessment, accreditation rarely incorporates outcome data, and feedback seldom informs programme improvement.

Consequently, quality assurance functions as fragmented rather than a continuous system, limiting validated educational outcomes. This finding is consistent with the CBME literature emphasising the need for alignment across competency definition, assessment, accreditation and feedback (4,21,22).

6 The “Invisible Curriculum–Unmeasured Outcomes” Gap

The findings indicate that much competence development in family practice remains invisible, unmeasured, and weakly governed. Although curricula define explicit competencies, key abilities, including contextual decision-making, longitudinal care, management of uncertainty, and professional identity, develop largely through the hidden curriculum (18,30,36). These complex, context-dependent competencies are inadequately captured by conventional assessment systems, creating an “invisible curriculum, unmeasured outcomes” gap in which the most important aspects of practice remain least measurable (20). This represents a fundamental limitation of current quality assurance models and highlights the need for more integrated, context-sensitive evaluation approaches.

Discussion

The findings show that the main challenge in family practice training is the fragmented integration of quality assurance components rather than their absence. Although competencies, assessment, accreditation, and feedback are well established, they remain poorly aligned. Competencies are not consistently translated into measurable outcomes, assessment lacks standardisation, accreditation emphasises institutional structures over educational effectiveness, and feedback is weakly linked to continuous improvement. Consequently, quality assurance is structurally established but functionally incomplete.

This fragmentation creates the “**invisible curriculum–unmeasured outcomes gap**,” whereby key competencies, such as contextual decision-making, longitudinal care, and professional identity, remain insufficiently evaluated. Although CBME frameworks, including WONCA, have strengthened the multidimensional conceptualisation of competence, encompassing technical, relational, ethical, and community-oriented domains (4,7,8), their operationalisation into valid, standardised, and measurable outcomes remains incomplete (11,22).

Operational Gap in Competency-Based Frameworks

Although CBME has clarified the multidimensional nature of competence in primary care, competencies remain insufficiently operationalised into measurable and standardised outcomes (4,7,11,20–23). Consequently, competency frameworks often function as normative goals rather than empirically verifiable constructs, underscoring the need for programmatic assessment approaches that link competencies with reproducible educational outcomes (11,20,34).

Assessment Systems: Multiplicity without Coherence

The adoption of OSCEs, portfolios, and workplace-based assessments reflects a shift toward multidimensional evaluation (7,20,24). However, variability in implementation, scoring, assessor judgement, and programme design limits reliability, comparability, and longitudinal evaluation of competence (14,20,21). Thus, multiple assessment tools alone do not ensure meaningful competency measurement or system-level quality assurance (20,21,34).

Accreditation Systems: Structural Strength, Limited Outcome Evidence

Accreditation frameworks provide governance, accountability, and institutional standardisation (28), but primarily assess structures and processes rather than educational outcomes (29). In family practice, where competence is relational, longitudinal, and context-dependent (30), accreditation alone cannot demonstrate graduates' competence. Integrating validated outcome measures into accreditation is therefore essential (22).

Feedback Mechanisms: Data without System Learning

Although feedback mechanisms are widely implemented (16), they rely largely on learner perceptions rather than objective competence measures (17). Feedback rarely informs curriculum redesign or institutional decision-making, reflecting the absence of closed-loop quality improvement and limiting adaptive system learning (16,58).

Cross-Level Misalignment

Competencies, assessment, accreditation, and feedback operate at different system levels but remain weakly integrated. This fragmentation disconnects competency definition, measurement, validation, and improvement, limiting verified educational outcomes and reinforcing concerns raised in the CBME literature (4,21,22).

The Hidden Curriculum

Much competence in family practice develops through experiential learning, longitudinal patient care, and management of uncertainty (18,30). Because current quality assurance prioritises observable and standardised indicators, these contextual and relational competencies remain underrepresented, highlighting the need for more context-sensitive evaluation models (20).

Digital Transformation

Artificial intelligence, telemedicine, and data-driven care require new competencies in digital literacy, algorithm interpretation, and adaptive decision-making (23,25–27). However, these competencies remain insufficiently integrated into existing curricula and assessment systems, widening the gap between training and clinical practice.

Toward an Integrated Quality Assurance System

The principal limitation of current quality assurance is the lack of integration across its components. Effective systems require alignment of competency frameworks, assessment, accreditation, and feedback within a continuous, outcome-oriented cycle (4,21).

Competencies should be translated into measurable outcomes, assessment should generate comparable longitudinal evidence, accreditation should incorporate outcome indicators, and feedback should drive continuous improvement. Such a dynamic, system-level approach provides a stronger foundation for competency-based education and adaptation to future challenges.

What did we Learn?

This study demonstrates that the principal challenge in family practice training is structural rather than conceptual. Competency frameworks are well established, but their translation into measurable outcomes remains incomplete. Assessment, accreditation, and feedback operate as fragmented rather than integrated processes, limiting the measurement, validation, and continuous improvement of competence.

The key lesson is that quality depends on system-level integration rather than stronger individual components. Competencies, assessment, accreditation, and feedback should function within a unified, feedback-driven framework that captures both measurable outcomes and the contextual, longitudinal, and professional dimensions of family practice.

Strengthening quality therefore requires integrating existing frameworks into a coherent, adaptive system rather than creating new ones.

Conclusion

Family practice training is central to both medical education and health system performance, yet its quality assurance remains fragmented. Although competencies, assessment, accreditation, and feedback are well established, they operate as poorly integrated components, limiting measurable and comparable outcomes. Competencies are defined but insufficiently operationalised, assessment lacks standardisation, accreditation focuses on structures rather than outcomes, and feedback is weakly linked to continuous improvement.

Key competencies developed through the hidden curriculum, including contextual decision-making, longitudinal care, and professional identity, remain inadequately evaluated. Digital transformation further widens the gap between training and practice by introducing competencies not yet systematically addressed. Improving family practice training therefore requires reconceptualising quality assurance as an integrated, feedback-driven system in which competencies become not only defined but also measurable, validated, and continuously improved.

Acknowledgement: no funding received.

References

- Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q.* 2005;83(3):457-502. doi: 10.1111/j.1468-0009.2005.00409.x. PMID: 16202000; PMCID: PMC2690145.
- Kringos DS, Boerma GW, Hutchinson A, Saltman RB, editors. *Building primary care in a changing Europe* [Internet]. Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2015. PMID: 29035488.
- World Health Organization; United Nations Children's Fund (UNICEF). *Primary health care: transforming vision into action: operational framework*. Geneva: World Health Organization; 2018. Available from: <https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.16>
- Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet.* 2010 Dec 4;376(9756):1923-58. doi: 10.1016/S0140-6736(10)61854-5. Epub 2010 Nov 26. PMID: 21112623.
- Organisation for Economic Co-operation and Development (OECD). *Realising the potential of primary health care*. Paris: OECD Publishing; 2020. Available from: https://www.oecd.org/en/publications/realising-the-potential-of-primary-health-care_a92ade4-en.html
- Harden RM. Outcome-Based Education: the future is today. *Med Teach.* 2007 Sep;29(7):625-9. doi: 10.1080/01421590701729930. PMID: 18236247.
- Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver I, Taber S, Talbot M, Harris KA. Competency-based medical education: theory to practice. *Med Teach.* 2010;32(8):638-45. doi: 10.3109/0142159X.2010.501190. PMID: 20662574.
- WONCA Europe. The European definition of general practice/family medicine. 2011 [cited 2026 Apr 20]. Available from: <https://www.woncaeurope.org/page/definition-of-general-practice-family-medicine>
- World Health Organization. *Transforming and scaling up health professionals' education and training*. Geneva: World Health Organization; 2013. Available from: <https://www.who.int/publications/i/item/transforming-and-scaling-up-health-professionals-education-and-training>
- Allen J, Gay B, Crebolder H, Heyrman J, Svab I, Ram P. The European definitions of the key features of the discipline of general practice: the role of the GP and core competencies. *Br J Gen Pract.* 2002 Jun;52(479):526-7. PMID: 12051237; PMCID: PMC1314348.
- Ten Cate O. Competency-Based Postgraduate Medical Education: Past, Present and Future. *GMS J Med Educ.* 2017 Nov 15;34(5):Doc69. doi: 10.3205/zma001146. PMID: 29226237; PMCID: PMC5704607.
- Gruppen LD, Mangrulkar RS, Kolars JC. The promise of competency-based education in the health professions for improving global health. *Hum Resour Health.* 2012 Nov 16;10:43. doi: 10.1186/1478-4491-10-43. PMID: 23157696; PMCID: PMC3543172.
- van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005 Mar;39(3):309-17. doi: 10.1111/j.1365-2929.2005.02094.x. PMID: 15733167.
- Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ.* 2013 Dec;47(12):1164-74. doi: 10.1111/medu.12289. PMID: 24206150.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990 Sep;65(9 Suppl):S63-7. doi: 10.1097/00001888-199009000-00045. PMID: 2400509.
- Hattie J, Timperley H. The power of feedback. *Rev Educ Res.* 2007;77(1):81-112. doi:10.3102/003465430298487
- Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract.* 2012 Mar;17(1):15-26. doi: 10.1007/s10459-011-9290-7. Epub 2011 Apr 6. PMID: 21468778; PMCID: PMC3274671.
- Hafferty FW. Beyond curriculum reform: confronting medicine's hidden curriculum. *Acad Med.* 1998 Apr;73(4):403-7. doi: 10.1097/00001888-199804000-00013. PMID: 9580717.

19. Mahood SC. Medical education: Beware the hidden curriculum. *Can Fam Physician*. 2011 Sep;57(9):983-5. PMID: 21918135; PMCID: PMC3173411.
20. Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013 Jul;35(7):564-8. doi: 10.3109/0142159X.2013.789134. Epub 2013 Apr 30. PMID: 23631408.
21. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-14. doi: 10.3109/0142159X.2012.652239. PMID: 22364452.
22. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-82. doi: 10.3109/0142159X.2010.500704. PMID: 20662580.
23. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019 Jan;25(1):44-56. doi: 10.1038/s41591-018-0300-7. Epub 2019 Jan 7. PMID: 30617339.
24. Morley J, Machado CCV, Burr C, Cows J, Joshi I, Taddeo M, Floridi L. The ethics of AI in health care: A mapping review. *Soc Sci Med*. 2020 Sep;260:113172. doi: 10.1016/j.socscimed.2020.113172. Epub 2020 Jul 15. PMID: 32702587.
25. World Health Organization. Global strategy on digital health 2020–2025. Geneva: World Health Organization; 2021 [cited 2026 Apr 20]. Available from: <https://www.who.int/publications/i/item/9789240020924>
26. S Woods S, M Greene S, Adams L, Cordovano G, F Hudson M. From E-Patients to AI Patients: The Tidal Wave Empowering Patients, Redefining Clinical Relationships, and Transforming Care. *J Particip Med*. 2025 May 16;17:e75794. doi: 10.2196/75794. PMID: 40378413; PMCID: PMC12101788.
27. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *J Med Internet Res*. 2006 Jun 16;8(2):e9. doi: 10.2196/jmir.8.2.e9. PMID: 16867972; PMCID: PMC1550701.
28. European Association for Quality Assurance in Higher Education (ENQA); European Students' Union (ESU); European University Association (EUA); European Association of Institutions in Higher Education (EURASHE). Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG 2015). Brussels: ENQA; 2015.
29. Harvey, L.; Williams, J. Fifteen Years of Quality in Higher Education. *Qual. High. Educ*. 2010, 16, 3–36, <https://doi.org/10.1080/13538321003679457>.
30. Eraut M. Developing professional knowledge and competence. London: Falmer Press; 1994.
31. Walt G, Gilson L. Reforming the health sector in developing countries: the central role of policy analysis. *Health Policy Plan*. 1994 Dec;9(4):353-70. doi: 10.1093/heapol/9.4.353. PMID: 10139469.
32. Richard G. Roberts, Vincent R. Hunt, Teresa I. Kulie, Wesley Schmidt, Julie M. Schirmer, Tiago Villanueva, C. Ruth Wilson. Family medicine training—the international experience. *Med J Aust*. 2011;194(11):S84–7. doi:10.5694/j.1326-5377.2011.tb03135.x
33. Michels NRM, Maagaard R, Buchanan J, Scherpbier N. Educational training requirements for general practice/family medicine specialty training: recommendations for trainees, trainers and training institutions. *Educ Prim Care*. 2018 Nov;29(6):322-326. doi: 10.1080/14739879.2018.1517391. Epub 2018 Sep 12. PMID: 30207897.
34. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007 Nov;29(9):855-71. doi: 10.1080/01421590701775453. PMID: 18158655.
35. David A. Kolb. *Experiential learning: experience as the source of learning and development*. Englewood Cliffs (NJ): Prentice Hall; 1984.
36. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. Reframing medical education to support professional identity formation. *Acad Med*. 2014 Nov;89(11):1446-51. doi: 10.1097/ACM.0000000000000427. PMID: 25054423.
37. David J. Nicol, Debra Macfarlane-Dick. *Formative assessment and self-regulated learning: a model and seven principles of good feedback practice*. Stud High Educ. 2006;31(2):199–218. doi:10.1080/03075070600572090.
38. Joint Commission on Accreditation of Healthcare Organizations. *Improvement: a guide to improving your patients' care*. 2nd ed. Oakbrook Terrace (IL): Joint Commission Resources; 2012.
39. World Health Organization. Global strategy on digital health 2020–2025. Geneva: World Health Organization; 2021. Available from: WHO publication page
40. Atun R, Aydın S, Chakraborty S, Sümer S, Aran M, Gürol I, Nazlıoğlu S, Özgülcü S, Aydoğan U, Ayar B, Dilmen U, Akdağ R. Universal health coverage in Turkey: enhancement of equity. *Lancet*. 2013 Jul 6;382(9886):65-99. doi: 10.1016/S0140-6736(13)61051-X. Epub 2013 Jun 27. PMID: 23810020.
41. Tatar M, Mollahaliloğlu S, Sahin B, Aydın S, Maresso A, Hernández-Quevedo C. Turkey. Health system review. *Health Syst Transit*. 2011;13(6):1-186, xiii-xiv. PMID: 22455830.
42. Recep Akdağ. Turkey Health Transformation Program: progress report. Ankara: Ministry of Health; 2012.
43. T.C. Sağlık Bakanlığı Tıpta Uzmanlık Kurulu. Tıpta ve dış hekimliğinde uzmanlık dallarının çekirdek müfredat ve standartları v.2.4 dokümanları (TUKMOS) [Internet]. Ankara: T.C. Sağlık Bakanlığı; [cited 2026 Apr 21]. Available from: <https://tuk.saglik.gov.tr/TR-84694/v-24-dokumanlari.html>
44. Güneş ED, Yaman H. Transition to family practice in Turkey. *J Contin Educ Health Prof*. 2008 Spring;28(2):106-12. doi: 10.1002/chp.167. PMID: 18521878.
45. Yükseköğretim Kalite Kurulu (YÖKAK). Kurumsal Akreditasyon Programı Rehberi. Ankara: Yükseköğretim Kalite Kurulu; 2021.
46. World Organization of Family Doctors (WONCA). *WONCA global standards for postgraduate family medicine education*. Bangkok: WONCA; 2013.

47. European Union. Directive 2005/36/EC of the European Parliament and of the Council of 7 September 2005 on the recognition of professional qualifications. Off J Eur Union. 2005;L255:22–142.
48. T.C. Sağlık Bakanlığı. Sağlıkta Dönüşüm Programı [Internet]. Ankara: T.C. Sağlık Bakanlığı; [cited 2026 Apr 22]. Available from: <https://www.saglik.gov.tr/TR,11415/saglikta-donusum-programi.html>
49. T.C. Sağlık Bakanlığı. Eğitim ve hizmet sunumuna ilişkin resmi dokümanlar [Internet]. Ankara: T.C. Sağlık Bakanlığı; [cited 2026 Apr 22]. Available from <https://www.saglik.gov.tr>
50. Akdeniz University Faculty of Medicine. Post-graduate education programme [Internet]. Antalya: Akdeniz University; [cited 2026 Apr 22]. Available from: https://tip.akdeniz.edu.tr/en/postgraduate_education_programme-10199
51. Istanbul University Faculty of Medicine. Postgraduate medical education documentation [Internet]. Istanbul: Istanbul University; [cited 2026 Apr 22]. Available from: <https://istanbul.edu.tr>
52. Ege University Faculty of Medicine. Specialty training [Internet]. Izmir: Ege University; [cited 2026 Apr 22]. Available from: https://med.ege.edu.tr/eng-2088/specialty_training.html
53. Ankara University Faculty of Medicine. Specialist residency programs [Internet]. Ankara: Ankara University; [cited 2026 Apr 22]. Available from: <https://www.medicine.ankara.edu.tr/en/specialist-residency-programs/>
54. Dokuz Eylül University Faculty of Medicine. Faculty of Medicine [Internet]. Izmir: Dokuz Eylül University; [cited 2026 Apr 22]. Available from: <https://www.medicine.ankara.edu.tr/en/specialist-residency-programs/>
55. Atatürk University Faculty of Medicine. Training and evaluation framework [Internet]. Erzurum: Atatürk University; [cited 2026 Apr 22]. Available from: <https://atauni.edu.tr>
56. Yükseköğretim Kalite Kurulu (YÖKAK). Kurum iç değerlendirme raporu (KİDR) hazırlama kılavuzu (Sürüm 3.2.1) [Internet]. Ankara: Yükseköğretim Kalite Kurulu; 2025 [cited 2026 Apr 22]. Available from: https://www.yokak.gov.tr/documents/national-doc/KIDR_Haz%C4%B1rlama_K%C4%B1lavuzu_3.2.1_yeni.pdf?utm_source=chatgpt.com
57. Yükseköğretim Kalite Kurulu (YÖKAK). Institutional accreditation reports [Internet]. Ankara: Yükseköğretim Kalite Kurulu; [cited 2026 Apr 22]. Available from: <https://www.yokak.gov.tr>
58. Batalden PB, Davidoff F. What is “quality improvement” and how can it transform healthcare? Qual Saf Health Care. 2007 Feb;16(1):2-3. doi: 10.1136/qshc.2006.022046. PMID: 17301192; PMCID: PMC2464920.
59. Virginia Braun, Victoria Clarke. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. doi:10.1191/1478088706qp0630a.
60. Matthew B. Miles, A. Michael Huberman, Johnny Saldaña. Qualitative data analysis: a methods sourcebook. 3rd ed. Thousand Oaks (CA): Sage; 2014.
61. Judith Green, Nicki Thorogood. Qualitative methods for health research. 4th ed. Thousand Oaks (CA): Sage; 2018.