Spraakgestuurd rapporteren binnen de huisartsenzorg

Technische rapportage

Procesleider: Roald van Leeuwen

Auteurs: Katharina Abraham, Wendela de Lange

Onderzoekers: Katharina Abraham (KA), Wendela de Lange (WdL)

In opdracht van digizo.nu

Versie 1.0
21 januari 2026

Inhoud

1 Leeswijzer **3**

2 Onderzoeksdoel **4**

3 Methode **5**
3.1 Zoekstrategie en screening 5
3.2 Inclusie- en exclusiecriteria 5
3.3 Data extractie 6
3.4 Data synthese 6
3.5 Kwaliteitsbeoordeling 6

4 Resultaten **7**
4.1 Geïncludeerde literatuur 7
4.2 Samenvatting bevindingen 9

5 Discussie **22**

6 Conclusie **23**

7 Referenties **24**

Bijlagen **26**
Bijlage A – Meetplan 26
Bijlage B – Criteria grijze literatuur 27
Bijlage C – Zoekstrategie 28
Bijlage D – Tabel studies per meetplanindicator 34
Bijlage E – Evidence gap 35
Bijlage F – Toepassingen voor de HAZ 40

1 Leeswijzer

In deze technische rapportage zijn de stappen van het waardebepalend onderzoek vastgelegd. Deze rapportage kan niet worden beschouwd als een wetenschappelijk artikel maar volgt wel de standaardstructuur van een wetenschappelijke publicatie.

Allereerst zijn het onderzoeksdoel en de methoden beschreven, die inzicht geven in de uitgevoerde onderzoeksstappen volgens de methodiek Waardebepaling van Digizo.

In de resultatensectie zijn de geïncludeerde studies opgenomen en wordt een korte samenvatting gepresenteerd om een overzicht te geven van de resultaten per meetplanindicator. Daarnaast is een uitgebreide beschrijving opgenomen van de bevindingen met betrekking tot alle must-have en should-have indicatoren. Voorafgaand aan iedere uitgebreide beschrijving van de bevindingen is een samenvatting opgenomen. Alle geëxtraheerde data zijn beschikbaar in het Excel-bestand, desgewenst op te vragen.

De beoordeling van de methodologische kwaliteit van de studies volgt aansluitend op de resultatensectie. In de discussie worden de kwaliteitsbeoordeling van de studies, de beschikbaarheid van vergelijkingen tussen interventie- en controlegroepen, en relevante studiekenmerken (zoals zorgsetting en type interventieapplicatie) behandeld.

De rapportage eindigt met de discussie en conclusie van het onderzoek. Daarnaast is in Bijlage E een overzicht opgenomen van de ‘evidence gaps’ per meetplanindicator: een inschatting van de mate waarin er, op basis van het meetplan, voldoende bewijs beschikbaar is. Deze inschatting is gebaseerd op de beschikbaarheid van studies, het studiedesign, de methodologische kwaliteit en de steekproefgrootte.

2 Onderzoeksdoel

Het doel van dit onderzoek was het evalueren van een nieuwe werkwijze met spraak-gestuurd rapporteren (SGR) tijdens patiëntconsulten in de huisartsenzorg (HAZ), met betrekking tot kwaliteit, impact op de administratieve last, werktevredenheid, gebruik, patiënt-zorgprofessionalbetrokkenheid, houding van patiënten en zorgprofessionals ten aanzien van het gebruik van SGR, en kosten.

3 Methode

De methodologische aanpak van dit onderzoek is gebaseerd op de methodiek waardebepaling van Digizo.

3.1 Zoekstrategie en screening

De zoekstrategie is op 13 juli 2025 uitgevoerd door een informatiespecialist en richtte zich op spraakgestuurde rapportage (SGR) binnen de domeinen Sociale Domein, Verpleeg-, Verzorgingshuizen en Thuiszorg, Medische Specialisten Zorg en HAZ, zonder beperkingen ten aanzien van studiedesigns (zie Bijlage A voor de volledige zoekstrategie en databases).

De titel- en abstractscreening is uitgevoerd in Rayyan.ai door één onderzoeker (KA). Deze keuze is gemaakt omdat de zoekstrategie domein overstijgend was, waardoor het niet efficiënt was om een tweede onderzoeker als reviewer te betrekken, gezien hun beperkte inzicht in de overige domeinen. Na de selectie van relevante studies voor het domein HAZ is de full-tekstscreening uitgevoerd door twee onderzoekers (KA, WdL).

Ter controle van de kwaliteit van de zoekstrategie en screening is een kwaliteitscheck uitgevoerd op basis van één sleutelartikel (DOI: 10.2196/60020). Indien de zoekstrategie en screening correct zijn uitgevoerd, zou dit artikel geselecteerd moeten worden in zowel de titel- en abstractscreeneing als in de full-tekstbeoordeling.

3.2 Inclusie- en exclusiecriteria

De onderstaande tabel geeft weer welke populaties, interventies, vergelijkingen en uitkomsten (PICO) in dit onderzoek zijn meegenomen. Studies moesten in full-tekst beschikbaar zijn en in het Engels of Nederlands geschreven zijn. Er is in eerste instantie gekeken naar recente systematische reviews; indien deze aansloten bij de PICO, werden alleen de primaire studies die buiten de zoekperiode van deze reviews vielen, in full-tekst gescreend. Studies met een van de volgende studietypen zijn niet meegenomen in het onderzoek: scoping reviews, commentaren, editorials, essays, nieuwsitems, opinieartikelen en research letters.

De grijze literatuur is niet systematisch verzameld, maar op basis van een oproep aan (regionale) huisartsenorganisaties om bewijs voor de relevante uitkomsten aan te leveren. Daarnaast zijn tijdschriften zoals Huisarts & Wetenschap (H&W) en het Nederlandse Tijdschrift voor geneeskunde (NTVG) bekeken voor SGR. Voor de grijze literatuur zijn aanvullende inclusie-exclusiecriteria toegepast (zie Bijlage B).

PICO	Beschrijving
Populatie	Zorgprofessionals en patiënten in de huisartsenzorg
Interventie	Speech-to-Summary: Het consult wordt door de applicatie opgenomen, letterlijk (verbatim) uitgeschreven, vervolgens samengevat en gestructureerd volgens de SOEP-systematiek en opgeslagen in het Huisartsen Informatie Systeem (HIS).

Vergelijking	Reguliere verslaglegging Opmerking: Ook gegevens die uitsluitend betrekking hebben op de SOLL-situatie worden meegenomen in het onderzoek.
Uitkomsten *	• Kwaliteit van klinische samenvattingen • Houding patiënten • Houding zorgverleners • Gebruik • Werkplezier • Administratieve tijd • Tevredenheid met communicatie patiënten-zorgverlener (perspectief patiënt) • Tevredenheid met communicatie patiënten-zorgverlener (perspectief zorgverlener) • Toegankelijkheid

*De uitkomsten zijn in detail beschreven in het meetplan, te vinden in Bijlage A.

3.3 Data extractie

De data-extractie is uitgevoerd door één onderzoeker (KA). De studiekenmerken zijn geëxtraheerd en de relevante uitkomsten zijn vastgelegd met als uitgangspunt het meetplan (Bijlage A).
Alle geëxtraheerde data zijn beschikbaar in het Excel-bestand, desgewenst op te vragen.

3.4 Data synthese

Vanwege de heterogeniteit in studiedesigns is gekozen voor een narratieve samenvatting.
De uitkomsten zijn samengevat per meetplan indicator (Bijlage A).

3.5 Kwaliteitsbeoordeling

Primaire studies worden beoordeeld met de MMAT (zie kader 1), aangezien verschillende studiedesigns worden geïncludeerd.

MMAT
De Mixed Methods Appraisal Tool (MMAT) is een gevalideerd instrument dat wordt gebruikt voor de kwaliteitsbeoordeling van wetenschappelijke studies met uiteenlopende onderzoeksdesigns, waaronder kwalitatieve, kwantitatieve en mixed-methods studies. Het unieke van de MMAT is dat het binnen één raamwerk toelaat om verschillende typen studies systematisch en op een consistente manier te evalueren.
Het instrument bestaat uit vijf specifieke beoordelingscriteria per studiedesign, die zich richten op aspecten zoals de helderheid van de onderzoeksvraag, geschiktheid van het design, de kwaliteit van de dataverzameling, de analysemethoden en de interpretatie van de resultaten.
De MMAT wordt vooral toegepast in systematische reviews waarin verschillende soorten studies worden geïncludeerd. Het doel is om de methodologische kwaliteit van de afzonderlijke studies op een transparante en reproduceerbare manier in kaart te brengen, zodat de betrouwbaarheid van de conclusies uit de review beter kan worden ingeschat.

Kader 1

4 Resultaten

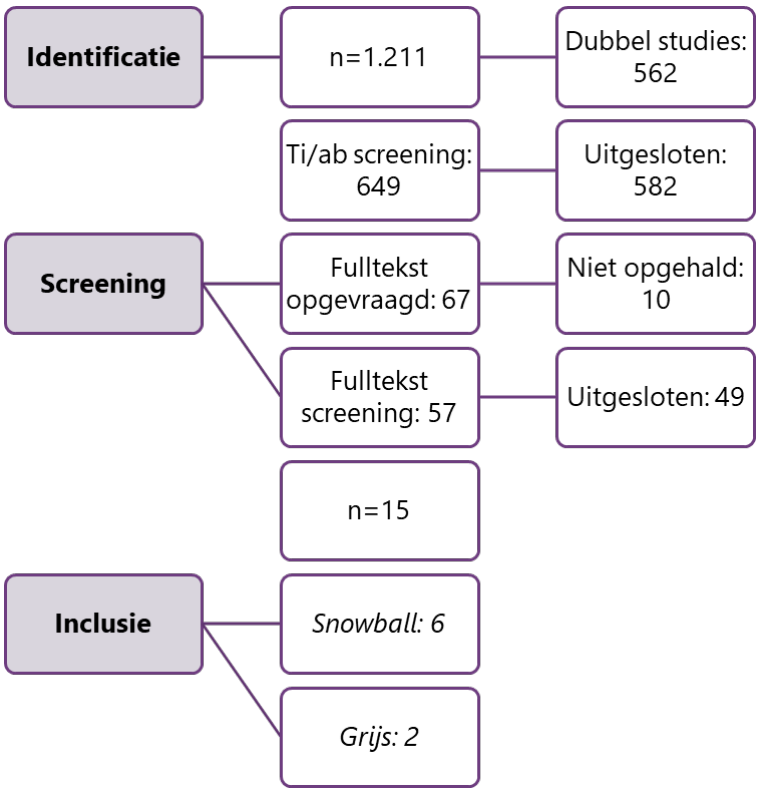
4.1 Geïncludeerde literatuur

Er zijn 1.211 records geïdentificeerd op basis van de zoekstrategie. Na het ontdubbelen zijn er 649 studies overgebleven en op basis van titel en abstract zijn deze gescreend op inclusie- en exclusiecriteria.

67 studies zijn geïncludeerd voor de HAZ voor full-tekst review (zie figuur onderaan) en 59 studies zijn geëxcludeerd vanwege:

- Full-tekst niet vindbaar (n=10)
- Studie design (n=14)
- De populatie (n=12),
- De interventie (n=20)
- De uitkomsten niet relevant (n=3)

Er zijn in totaal 15 studies gevonden, waarvan 6 via referentiescreening (snowballing) (zie figuur 1). Eén van de studies is een preprint, waarbij dus nog geen peerreview heeft plaatsgevonden. Deze werd als grijze literatuur meegenomen. Daarnaast is er nog een Nederland-specifieke pilotstudies als grijze literatuur geïncludeerd.



Figuur 1

Het design van de geïncludeerde studies is te vinden in de onderstaande tabel.

De geïncludeerde studies zijn voornamelijk gepubliceerd in 2024 en 2025 met één studie gepubliceerd in 2021. In acht studies betrof het de Amerikaanse zorgsetting (n=8/14; 57%), maar ook in Canada (n=1) en Nederland (n=2). In vier studies werd het land van de zorgsetting niet gerapporteerd. In zes studies betrof de zorgsetting een mix van huisartsenzorg en medisch-specialistische zorg: twee simulatiestudies met als doel de foutgevoeligheid van AI-gegenereerde notities te evalueren (Biro, Handley, Cobb, et al., 2025; Biro, Handley, Mickler, et al., 2025), drie studies waarin de populatie voornamelijk uit de huisartsenzorg kwam (>50%) (Ma et al., 2025; Shah, Crowell, et al., 2025; Shah, Devon-Sand, et al., 2025), en één studie waarin de verdeling niet bekend was (Liu et al., 2024).

Aantal	Design	Referentie(s)
1	Niet-gerandomiseerde klinische studie met pre-postvergelijking	(Liu et al., 2024)
1	Observationele studie met pre-postvergelijking	(Misurac et al., 2025)
1	Observationele studie met open- en geblindeerde fase met vergelijking tussen SGR en typen	(Owens et al., 2024)
2	Pre-post kwaliteitsverbeteringsstudies	(Ma et al., 2025; Shah, Devon-Sand, et al., 2025†)
2	Kwalitatieve studies zonder vergelijking	(Bundy et al., 2024; Shah, Crowell, et al., 2025‡)
7	Simulatiestudies: 1 met aanvullende mixed-methods pre-postvergelijking (Ha, 2024) 1 preprint (Anderson et al., 2025)	(Anderson et al., 2025; Biro, Handley, Cobb, et al., 2025; Biro, Handley, Mickler, et al., 2025; Hose et al., 2025; Van Buchem et al.*, 2024; Wang et al., 2021)
1	Pre- post pilot evaluatie van een implementatietraject	(Haak et al., 2025)

*Sleutelartikel †in-tekst referentie Shah_a ‡in-tekst referentie Shah_b

Kwaliteit van de studies

Alle wetenschappelijke studies zijn primaire studies en geëvalueerd met de MMAT. Zeven van deze studies (50%) werden door de beoordelaar als voldoende van kwaliteit beschouwd, aangezien zij op meer dan 50% van de MMAT-vragen over kwaliteit positief scoorden (zie de tabel). In sommige studies was het antwoord op een vraag niet beschreven of onduidelijk, deze is dan als 'negatief' gescoord. Onvoldoende kwaliteit werd vooral gevonden in de simulatiestudies maar ook in de observationele studies. Hoewel sampling belangrijk zou zijn om de onderzoeksvraag (kwaliteit van AI-gegenereerde notities) te beantwoorden, is er gekozen voor een simulatiedesign met óf mockconsulten óf real-worldconsulten waarbij niet te beoordelen is hoe de selectie is uitgevoerd of voor welke medische specialisatie. Ook de selectie van beoordelaars van de AI-notities was onduidelijk of niet representatief (bijvoorbeeld studenten). Daardoor is de steekproef als geheel niet representatief. In één van de observationele studies (Misurac) was geen fulltekst beschikbaar om de kwaliteit van de studie te beoordelen; ook als was het wel mogelijk om de relevante data te extraheren. Deze studie is ook beoordeeld als van onvoldoende kwaliteit, gezien het feit dat het niet mogelijk was om te dit te beoordelen.

(Eerste) auteur	Studiedesign	Kwaliteitsbeoordeling*
Ha, et al.	Mixed-methodsstudie	Voldoende
Shah, Crowell, et al.	Kwalitatieve studie	Voldoende
Bundy, et al.	Kwalitatieve studie	Voldoende
Liu, et al.	Niet-gerandomiseerd gecontroleerd studie	Voldoende
Misurac, et al.	Observationele studie	Onvoldoende
Ma, et al.	Observationele studie	Onvoldoende
Owens, et al.	Observationele studie	Voldoende
Van Buchem, et al.	Simulatiestudie	Onvoldoende
Hose, et al.	Simulatiestudie	Onvoldoende
Anderson, et al.	Simulatiestudie	Onvoldoende
Biro, Handley, Cobb, et al.	Simulatiestudie	Voldoende
Biro, Handely, Mickler, et al.	Simulatiestudie	Voldoende
Wang, et al.	Simulatiestudie	Onvoldoende
Shah, Devon-Sand, et al.	Kwantitatieve studie	Onvoldoende

* Studies met >50% positieve antwoorden op de kwaliteitsvragen werden als voldoende van kwaliteit beschouwd.

4.2 Samenvatting bevindingen

Overzicht

Must-have 1: Veiligheid en kwaliteit van de samenvatting

Er bestaan gemengde meningen onder zorgverleners over de kwaliteit van AI-gegenereerde notities: soms worden deze als gelijkwaardig of zelfs beter beschouwd dan handmatige verslaglegging. Voor onbewerkte notities is echter een significant verschil gevonden ten nadele van AI, terwijl gecorrigeerde notities vergelijkbaar scoren. Meerdere onderzoeken wijzen op beperkingen in nauwkeurigheid en betrouwbaarheid, met name wat betreft volledigheid en bondigheid. De grijze literatuur suggereert daarnaast dat er kwaliteitsverschillen kunnen bestaan tussen verschillende AI-spraaktoepassingen, maar ook hier bereiken de kwaliteitsscores een niveau dat vergelijkbaar is met handmatig opgestelde notities.

Must-have 2: Acceptatie door patiënten

Patiënten hebben sterk uiteenlopende gevoelens over AI-spraak toepassing en privacy. Aan de ene kant willen sommige patiënten controle behouden en de mogelijkheid hebben hun toestemming in te trekken. Aan de andere kant vinden sommige patiënten privacy minder relevant. Hun focus ligt op het gezondheidsproblemen en zij ervaren voldoende vertrouwen te hebben in de zorgprofessional.

Must-have 3: Adoptie zorgprofessional

Het gebruik van AI-spraak applicaties na implementatie ligt bij consulten tussen de 50% en 85% en varieert sterk tussen zorgprofessionals.

Zorgprofessionals hebben een gemengd beeld van de bruikbaarheid van AI-spraaktoepassingen. Het gebruik wordt vaak als eenvoudig ervaren, maar niet in alle situaties als toepasbaar beschouwd, bijvoorbeeld bij korte of sociale consulten en bij niet-Engelssprekende patiënten. Wel geven zorgprofessionals aan dat het, na afweging van voor- en nadelen, de moeite waard is. Verder zijn zorgprofessionals over het algemeen positief over langdurig gebruik van AI-spraaktoepassingen en over het verwachte toekomstige gebruik door collega's.

Should-have 1: Werkplezier

AI-spraakapplicaties dragen bij aan minder burn-out en een lagere mentale/cognitieve belasting, met een verbeterd werkplezier voor zorgprofessionals.

Should-have 2: Administratieve werkdruk

AI-spraaktoepassingen verlagen de tijd die nodig is voor verslaglegging aanzienlijk, ook buiten werktijd. Het bewerken van AI-gegenereerde notities kost volgens zorgprofessionals een matige inspanning. AI-spraaktoepassingen verschillen mogelijk onderling in de mate van tijdsbesparing.

Should-have 3: Tijdigheid – Aandacht voor het gesprek zorgprofessional perspectief

Zorgprofessionals gaven aan dat AI-spraakapplicaties hen hielpen zich beter te concentreren op het gesprek en meer aandacht aan de patiënt te besteden. De toepassingen werden gezien als ondersteunend bij het bevorderen van betrokkenheid tijdens consulten en bij het verbeteren van de kwaliteit van de interactie met patiënten.

Should-have 4: Tijdigheid – Aandacht voor het gesprek patiënt perspectief

Patiënten ervaren de patiënt-zorgprofessionalrelatie over het algemeen als persoonlijker en rapporteren dat er tijdens een consult met een AI-spraaktoepassing minder tijd aan typen wordt besteed. De algemene tevredenheid van patiënten is echter vergelijkbaar met die bij consulten zonder AI-spraaktoepassing.

Should-have 5: Toegankelijkheid

Er zijn geen studies gevonden.

Must-have 1 – Veiligheid en kwaliteit van de samenvatting

Belangrijke uitkomsten die voor deze must-have zijn gerapporteerd, betreffen de algemene kwaliteit, evenals specifieke kwaliteitsaspecten zoals accuraatheid, relevantie, correcte organisatie en juistheid (biases, fouten, 'hallucinaties'). De kwaliteitsuitkomsten zijn gemeten met de PDQI-9-vragenlijst*, op een Likertschaal of descriptief beschreven.

*De Physician Documentation Quality Instrument-9 (PDQI-9) is specifiek ontwikkeld voor elektronische medische documentatie, waaronder voortgangsverslagen en ontslagbrieven. De PDQI-9 beoordeelt negen kenmerken van verslaglegging: actualiteit, nauwkeurigheid, volledigheid, bruikbaarheid, organisatie, begrijpelijkheid, bondigheid, mate van synthese en interne consistentie. Elk kenmerk wordt gescoord op een vijfpuntsschaal. De som van de afzonderlijke scores resulteert in een totaalscore (maximaal 45) die de algemene kwaliteit van de documentatie weerspiegelt. Er bestaan geen officiële cut-offs binnen de PDQI-9.

Algemene kwaliteit – kwalitatief

Samenvattend: In het merendeel van de geïncludeerde kwalitatieve studies zijn de meningen over de kwaliteit van AI-gegenereerde notities positief: ze worden als gelijkwaardig of zelfs beter beschouwd dan handmatige verslaglegging. Een bron uit de Nederlandse setting laat zien dat zorgprofessionals in meerderheid (zeer) positief waren over de kwaliteit van de verslaglegging. Tegelijkertijd wijzen twee onderzoeken op beperkingen, zoals de volledigheid, opbouw en opmaak van de notities, en de bondigheid en nauwkeurigheid. Hierdoor blijft correctie en nabewerking noodzakelijk.

Zwarte literatuur

Vier studies rapporteerden over de kwaliteit van AI-gegenereerde notities (Van Buchem, Shah_a, Shah_b, Ha).

De studie van Van Buchem liet geneeskundestudenten de verslagkwaliteit beoordelen. Vier studenten rapporteerden dat de automatisch gegenereerde samenvattingen hun verwachtingen overtroffen, terwijl vier anderen de kwaliteit onvoldoende achtten vanwege de aanwezigheid van fouten en de extra tijd die nodig was voor correcties.

In de kwaliteitsverbeteringsstudie van Shah_a werden 46 respondenten bevraagd, waarvan 52% een verbeterde kwaliteit van klinische verslaglegging rapporteerde met AI-spraak applicaties.

In de kwalitatieve studie van Shah_b werden 19 zorgverleners bevraagd over hun mening van de kwaliteit van AI-spraak applicaties. Van hun opmerkingen was 58% positief, 11% neutraal en 32% negatief beschreven. Er werden vooral negatieve opmerkingen gemaakt over de volledigheid van de notities (76%), de opbouw van de notities (80%), de opmaak van de notities (70%) en de bondigheid van de notities (83%).

De mixed-methodsstudie van Ha schetst een positief beeld van de kwaliteit van AI-notities, waarbij zorgverleners rapporteerden dat deze een hogere kwaliteit en meer detail bevatten dan handmatig opgestelde notities.

Grijze literatuur

In de pre- post pilot evaluatie van Haak gebruiken 16 zorgverleners Juvoly en vulde 9 zorgverleners de evaluatievragenlijst, daarbij waren 6 van de 9 zorgverleners positief of zeer positief over de kwaliteit van de verslaglegging.

Algemene kwaliteit – kwantitatief

Samenvattend: De geïncludeerde kwantitatieve wetenschappelijke studies laten zien dat de kwaliteit van AI-gegenereerde notities vergelijkbaar wordt beoordeeld met die van niet-AI-gegenereerde notities. Voor onbewerkte notities is er echter een significant verschil gevonden ten nadele van AI, terwijl gecorrigeerde notities vergelijkbaar scoren met handmatige. De grijze literatuur suggereert daarnaast dat er kwaliteitsverschillen kunnen bestaan tussen verschillende AI-spraak applicaties voor spraakgestuurd rapporteren, maar ook hier bereiken de kwaliteitscores een niveau dat vergelijkbaar is met handmatig opgestelde notities.

Zwarte literatuur

Twee studies rapporteerden over de kwaliteit van AI-gegenereerde notities (Van Buchem, Shah_a).

In de simulatiestudie van Van Buchem werden de handmatig opgestelde notities (waarvan de best beoordeelde als referentie werd genomen) significant beter beoordeeld dan de onbewerkte AI-gegenereerde notities (Mediane PDQI-9 score: 31 vs. 25; p=<0,001). Wanneer de AI-gegenereerde notities echter door de gebruiker waren aangepast, werden deze vergelijkbaar beoordeeld met de beste handmatige notities (Mediane PDQI-9 score: 31 vs. 29; p=0,16). Voor de subcategorie bondigheid werden de handmatig opgestelde notities wel significant beter beoordeeld dan de bewerkte AI-gegenereerde notities (Mediane PDQI-9 score: 5 vs. 5; p=<0,001).

In de kwaliteitsverbeteringsstudie van Shah_a werd geen significant verschil gevonden in de kwaliteit van klinische verslaglegging vóór en na de introductie van een AI-spraak applicatie (Likertschaal 1-5: 2,47 vs. 2,21; p=0,173).

Grijze literatuur

In de grijze literatuur vergeleek een studie verschillende AI-spraak applicaties met elkaar op kwaliteit (Anderson). In de simulatiestudie werden significante verschillen tussen de toepassingen gevonden (Mediane PDQI-9-scores: 32,3 tot 39,4, p=<0.05), met een gemiddelde totaalscore van 36 over alle AI-spraak applicaties.

Accuraatheid

Samenvattend: De wetenschappelijke studies laten zien dat de accuraatheid van handmatige verslaglegging en AI-gegenereerde verslaglegging mét bewerking door de zorgverlener vergelijkbaar of hoger is. Notities zonder bewerking zijn echter minder accuraat dan handmatige verslaglegging. Vanuit het perspectief van zorgverleners bestaat er een gemengd beeld over de accuraatheid van AI-notities.

Zwarte literatuur

Drie studies rapporteerden over de accuraatheid van AI-gegenereerde notities (Van Buchem, Shah_b, Ha).

In de simulatiestudie van Van Buchem werden de handmatig opgestelde notities (waarvan de best beoordeelde als referentie werd genomen) significant accurater beoordeeld dan de onbewerkte AI-gegenereerde notities (Mediane PDQI-9 score: 5 vs. 4; p=<0,001). Wanneer de AI-gegenereerde notities echter door de gebruiker waren aangepast, werden deze vergelijkbaar beoordeeld met de beste handmatige notities (Mediane PDQI-9 score: 5 vs. 5; p=<0,99).

In de kwalitatieve studie van Shah_b betrof 85% van de opmerkingen negatieve bevindingen over de accuraatheid.

In de mixed-methodsstudie van Ha gaf 65,8% van de zorgverleners aan het eens te zijn dat de notities accuraat waren.

Grijze literatuur

In de pre-post pilot evaluatie van Haak gebruiken 16 zorgverleners Juvoly en vulde 9 zorgverleners de evaluatievragenlijst. Over de mate van benodigde eigen correctie waren ze overwegend positief. Over de nauwkeurigheid van verslagen werd wisselt gedacht: zo noemden meerdere zorgverleners de kwaliteit van verslaglegging waarbij soms foutieve registraties werden opgemerkt, de samenvatting niet altijd klopten en correctie nodig waren.

Bruikbaarheid en gestructureerdheid

Samenvattend: De bruikbaarheid en georganiseerdheid van bewerkte AI-notities werd gelijk beoordeeld als de handmatige verslagen. Zonder bewerking scoorden AI-notities minder goed dan handmatige verslaglegging.

Een studie rapporteerde over de bruikbaarheid en gestructureerdheid van AI-gegenereerde notities (Van Buchem).

In de simulatiestudie werden de handmatig opgestelde notities (waarbij de best beoordeelde als referentie werd genomen) significant bruikbaarder beoordeeld dan de onbewerkte AI-gegenereerde notities (mediane PDQI-9-score: 5 vs. 4; $p < 0{,}001$). Dit gold ook voor de gestructureerdheid (mediane PDQI-9-score: 4 vs. 4; $p = 0{,}03$).

Wanneer de AI-gegenereerde notities echter door de gebruiker waren aangepast, werd geen verschil gevonden ten opzichte van de beste handmatige notities (mediane PDQI-9-score bruikbaarheid: 5 vs. 4; $p < 0{,}10$; en gestructureerdheid: 4 vs. 4; $p = 0{,}89$).

Juistheid

Samenvattend: De geïncludeerde wetenschappelijke en grijze literatuur laat zien dat AI-gegenereerde notities regelmatig onnauwkeurigheden bevatten, met name in de vorm van omissies (weglatingen van informatie die wél in het consult voorkomt maar níet in de notitie belandt). Hoewel de mate van fouten varieert tussen studies en toepassingen, blijkt uit meerdere bronnen dat in een aanzienlijk deel van de notities één of meerdere fouten aanwezig zijn, met foutpercentages die uiteenlopen van circa 15% tot bijna 50%. Bias wordt daarentegen slechts zelden gerapporteerd, terwijl ernstige hallucinaties of toevoegingen minder vaak voorkomen dan omissies, maar wel expliciet werden genoemd door zorgverleners met ervaring in het gebruik van AI-spraak applicaties.

Zwarte literatuur

Vijf studies (Biro_a, Van Buchem, Hose, Bundy, Ha) onderzochten de foutgevoeligheid van AI-spraak applicaties.

In de simulatiestudie van Biro_a, waarin twee toepassingen werden vergeleken, werden in 70% van de notities fouten gevonden, met een gemiddeld aantal van 2,9 fouten per notitie (SD 2,7). De verdeling van fouttypen verschilde: omissies kwamen het vaakst voor (83% bij Product A en 54% bij Product B), gevolgd door irrelevante of verkeerd geplaatste tekst (25% bij Product B versus 6% bij Product A). Toevoegingen en foutieve weergave werden minder vaak gerapporteerd (respectievelijk 4–11% en 6–10%). De verschillen tussen Product A en Product B waren significant.

De simulatiestudie van Hose rapporteerde in totaal 66 fouten bij 11 consulten, wat neerkomt op gemiddeld zes fouten per consult. Het merendeel betrof omissies (55 fouten, 83%), gevolgd door foutieve outputs (6%), inhoud die op zichzelf correct maar klinisch ongepast was (6%) en toevoegingen (5%).

In de simulatie van Van Buchem werd beschreven dat een veelvoorkomende fout het ontbreken van negatieve symptomen betrof, zoals het niet vermelden van de afwezigheid van kortademigheid. Daarnaast ondervonden gebruikers technische problemen, waarbij de tool bleef hangen of zeer traag werkte door beperkingen in de rekencapaciteit.

In de mixed-methodsstudie van Ha werden vijf van de zes AI-spraak applicaties beoordeeld als goed, één als slecht en geen enkele als excellent met betrekking tot fouten. Daarnaast was 49% van de zorgverleners het ermee eens dat de gemaakte fouten van geringe aard waren. (De andere

antwoordopties werden niet gerapporteerd). De fouten betroffen meestal een verwijderd woord of een omissiefout.

In de kwalitatieve studie van Bundy noemden zorgverleners een aantal fouten bij het gebruik van AI-spraak applicaties, zoals hallucinaties, misgendering en toevoegingen.

Grijze literatuur

In de simulatiestudie van Anderson (preprint) werden de belangrijkste elementen uit 14 gesimuleerde gesprekken geëxtraheerd en geëvalueerd in de AI-gegenereerde notities met behulp van een referentienotitie. Gemiddeld over de vijf verschillende toepassingen werden in 26,3% van de gevallen (variërend van 16,7% tot 46,1%) fouten aangetroffen. Het merendeel hiervan betrof omissies (gemiddeld 76,3% van alle fouten), aangevuld met kleinere aantallen fouten door toevoegingen of gedeeltelijk correcte informatie.

Must-have 2 – Acceptatie door patiënt

Belangrijke uitkomsten die voor deze must-have zijn gerapporteerd, betreffen de privacy tijdens consulten met AI-spraak applicaties. Deze zijn beschrijvend weergegeven.

Privacy

Samenvattend: Patiënten hebben sterk uiteenlopende gevoelens over AI-spraak applicaties en privacy. Aan de ene kant willen sommige patiënten controle behouden en de mogelijkheid hebben hun toestemming in te trekken. Aan de andere kant vinden sommige patiënten privacy minder relevant, omdat hun focus ligt op gezondheidsproblemen en zij voldoende vertrouwen hebben in de zorgverlener.

Zwarte literatuur

Twee primaire studies rapporteerden over privacyzorgen bij patiënten (Shah_b, Ha).

In de mixed-methodsstudie van Ha gaven sommige patiënten aan dat zij bij ieder consult opnieuw toestemming wenste te geven voor het gebruik van een AI-spraak applicatie. Andere patiënten richtten hun aandacht vooral op hun gezondheidsproblemen in plaats van na te denken over ‘informed consent’. Verder waren er nauwelijks zorgen over dataprivacy en -beveiliging, mede door de goede relatie en het vertrouwen in de zorgverlener. Sommige patiënten voelden zich echter ongemakkelijk wanneer het ging om gevoelige onderwerpen, zoals financiële zaken.

In de kwalitatieve studie van Shah_b rapporteerden zorgverleners tijdens de interviews dat enkele patiënten terughoudend waren met het gebruik van een AI-spraak applicatie tijdens hun consult. Daarnaast vermeldden zorgverleners dat sommige patiënten zich na het gebruik van een AI-spraak applicatie ongemakkelijk voelden en niet opnieuw van de AI-spraak applicatie gebruik wilden maken.

Must-have 3 – Adoptie door zorgprofessional

Belangrijke uitkomsten die voor deze must-have zijn gerapporteerd, betreffen het gebruik, de bruikbaarheid en het toekomstig gebruik. Deze zijn gemeten met behulp van een Likertschaal, de System Usability Score (SUS) (0-100 scala) en beschrijvende analyses.

Gebruik

Samenvattend: Het gebruik van AI-spraak applicaties ligt bij consulten tussen de 50% en 85% en varieert sterk tussen zorgprofessionals.

Zwarte literatuur

Drie primaire studies rapporteerden over het gebruik van AI-spraak applicaties (Shah_a, Ha, Ma).

In de kwaliteitsverbeteringsstudie van Shah_a gaf ongeveer 65% van de zorgverleners aan AI-spraak applicaties vaak te gebruiken voor verslaglegging.

In de kwaliteitsverbeteringsstudie van Ma lieten de gebruiksstatistieken zien dat AI-spraak applicaties werden toegepast in de helft van alle consulten. Er was een significant verschil tussen zorgverleners: sommige gebruikten de AI-spraak applicatie bij vrijwel alle consulten, terwijl anderen deze slechts minimaal inzetten.

In de mixed-methodsstudie van Ha werd geen significant verschil gevonden in het gebruik tijdens consulten tussen drie verschillende AI-spraak applicaties. Wel gaf 85% van de zorgverleners aan de AI-spraak applicatie dagelijks of meerdere keren per week te gebruiken.

Bruikbaarheid

Samenvattend: De wetenschappelijke studies laten een gemengd beeld zien van de bruikbaarheid van AI-spraak applicaties. Het gebruik werd vaak als eenvoudig ervaren, maar bleek niet in alle situaties toepasbaar, zoals bij korte of sociale consulten en bij niet-Engelssprekende patiënten. In één studie gaf meer dan tweederde van de zorgverleners aan dat het na een afweging tussen voordelen en nadelen de moeite waard was.

Zwarte literatuur

Vier primaire studies rapporteerden over de bruikbaarheid van AI-spraak applicaties (Shah_a, Ha , Bundy, Shah_b, Van Buchem).

In de pre-post kwaliteitsverbeteringsstudie van Shah_a werd een significant hogere bruikbaarheid van de applicatie gevonden (gemiddelde SUS-score 0–100: 58,09 vs. 69,01; $p = 0,003$) na introductie van de AI-spraakapplicatie. Bovendien was 98% van de zorgverleners het ermee eens dat de AI-spraakapplicatie eenvoudig te leren en te gebruiken was. Na de introductie van de AI-spraakapplicatie werd het leergemak bovendien significant beter beoordeeld (Likertschaal 1 = eens tot 5 = helemaal oneens: 2,08 vs. 1,47; $p = 0,001$).

In de mixed-methodsstudie van Ha gaf 73% van de zorgverleners aan dat de voordelen van AI-spraak applicaties opwogen tegen de nadelen.

In de kwalitatieve studies van Bundy en Shah_b werden beperkingen genoemd, zoals beperkte functionaliteit bij niet-Engelssprekende patiënten en het feit dat toegang voor zorgverleners afhankelijk was van specifieke apparaten. Verder vonden zorgverleners het gebruik van AI-spraak applicaties passend bij urgente problemen, maar minder toepasbaar bij sociale consulten of korte consulten van minder dan tien minuten. In de kwalitatieve studie van Bundy uitten zorgverleners zorgen dat de introductie van AI-spraak applicaties zou kunnen leiden tot beleid waarin zij meer patiënten zouden moeten zien.

In de Nederlandse simulatiestudie met mock-consultaties (Van Buchem) gaven studenten een gemengd beeld van het gebruik van AI-spraak applicaties. De helft van de 18 medische studenten gaf

aan dat de AI-spraak applicatie prettig was om te gebruiken, zes studenten vonden de tool gemakkelijk in gebruik, vier studenten geloofden in het potentieel van de tool en vier studenten vonden de geproduceerde samenvattingen beter dan verwacht. Daarentegen vonden vier studenten dat de kwaliteit van de samenvattingen slecht was.

Grijze literatuur

In de pre-post pilot evaluatie van Haak gebruiken 16 zorgverleners Juvoly, van de gebruikers die de evaluatievragenlijst invulde (n=9) zou 89% Juvoly aanraden aan collega's.

Toekomstig gebruik

Samenvattend: Het merendeel van de zorgprofessionals was positief over het langetermijn gebruik van AI-spraak applicaties en over het verwachte toekomstige gebruik door collega's.

Zwarte literatuur

Drie primaire studies rapporteerden over het toekomstige gebruik van AI-spraak applicaties (Shah_a, Ha, Van Buchem).

In de pre-post kwaliteitsverbeteringsstudie van Shah_a werd geen significante verandering gevonden in de verwachtingen van zorgverleners over het lange termijn gebruik van AI-spraak applicaties of over de mate waarin collega's geïnteresseerd zouden zijn in het gebruik ervan. Wel gaf 78% van de respondenten aan zelf bereid te zijn de AI-spraak applicatie ook op de lange termijn te gebruiken, en 80% was ervan overtuigd dat collega's dit in de toekomst eveneens zouden doen.

In de mixed-methodsstudie van Ha gaf 82,3% van de zorgverleners aan in de toekomst gebruik te willen blijven maken van de AI-spraak applicatie.

In de studie van Van Buchem wilde tweederde van de medische studenten de huidige AI-spraak applicaties in de toekomst gebruiken, terwijl een derde dit pas zou doen wanneer de AI-spraak applicatie verder verbeterd was.

Should-have 1 – Werkplezier

Belangrijke uitkomsten die voor deze should-have zijn gerapporteerd, betreffen 'burnout', 'professional fulfillment' en 'task load' (werkdruk – niet specifiek administratieve werkdruk). Deze zijn vooral kwalitatief gemeten, soms met behulp van een Likertschaal of de PFI-WE*.

*De PFI-WE (Professional Fulfillment Index – Work Exhaustion subscale) is een gevalideerd instrument dat is ontwikkeld om professionele vervulling en burn-out bij zorgprofessionals te meten op een schaal van 0 tot 10. De subschaal Work Exhaustion richt zich specifiek op het domein van werkgerelateerde uitputting, een kerncomponent van burn-out.

Burn-out

Samenvattend: De kwantitatieve wetenschappelijke studies laten zien dat AI-spraakapplicaties burn-out verlagen. Op basis van de kwalitatieve studies kan hierover geen uitspraak worden gedaan.

Zwarte literatuur

Vier studies geven inzicht in de invloed van AI-spraak applicaties op burn-out (Misurac, Shah_a, Bundy, Ha).

In de observationele studie van Misurac werd een 38% lagere burn-outprevalentie gerapporteerd na de implementatie van een AI-spraak applicatie (69% vs. 43%). Verder werd een significante verbetering gevonden in werkplezier en burn-out (Gemiddelde PFI-WE score: 4,16 vs. 3,16; p=0,005). De score lag bovendien onder de afkappunten voor burn-out (3,33).

In de kwaliteitsverbeteringsstudie van Shah_a werd na de introductie van een AI-spraak applicatie een significante reductie in burn-out gevonden (Gemiddelde PFI-WE score: 7,47 vs. 5,53; p<0,001).

In de mixed-methodsstudie van Ha gaf 55% van de zorgverleners aan dat AI-spraak applicaties burn-out verminderd.

In de kwalitatieve studie van Bundy maakte één van de twaalf zorgverleners een opmerking over burn-out, namelijk dat door het gebruik van AI-spraakapplicaties een positief effect op stress en slaap te verwachten was.

Werkplezier

Samenvattend: De gevonden wetenschappelijke en grijze studies suggereren dat AI-spraak applicaties leiden tot minder frustratie bij het gebruik van het HIS en tot een verbeterd of positief werkplezier.

Zwarte literatuur

Twee primaire studies rapporteerden over het werkplezier van zorgprofessionals (Ha, Liu).

In de mixed-methodsstudie van Ha gaf meer dan de helft (58,6%) van de zorgprofessionals aan een verbeterd werkplezier te ervaren met AI-spraak applicaties. (De andere antwoordopties werden niet gerapporteerd).

Liu rapporteerde in een niet-gerandomiseerde studie dat significant meer zorgprofessionals minder frustratie ervoeren bij het gebruik van het HIS na de introductie van AI-spraak applicaties (14,5% vs. 44,7%; p=0,001).

Grijze literatuur

Vanuit de pre post pilot evaluatie van Haak, waarbij 9 van de 16 zorgverleners (huisartsen en praktijkondersteuners) de evaluatie invulde, waren de ervaringen over de impact op het werkplezier overwegend positief bij gebruik van Juvoly.

Taakbelasting

Taakbelasting ('Task load') beschrijft de (ervaren) mentale en tijdsbelasting tijdens het werk.

Samenvattend: De wetenschappelijke studies laten een verbetering zien van AI-spraak applicaties op de mentale en tijdsbelasting van zorgverleners.

Zwarte literatuur

Vier primaire studies rapporteerden over de impact van AI-spraak applicaties op de taakbelasting van zorgverleners (Ha, Shah_a, Shah_b, Bundy).

In de mixed-methodsstudie van Ha rapporteerde 75,7% van de zorgverleners een reductie in cognitieve belasting na introductie van AI-notities.

In de kwaliteitsverbeteringsstudie van Shah_a werd een significante verbetering gevonden in taakbelasting (Gemiddelde Provider Task Load [PTL]-index, schaal 0–100: 68,98 vs. 44,56; p<0.001).

In de kwalitatieve studie van Shah_b waren, na introductie van AI-notities, alle opmerkingen over cognitieve/mentale belasting positief. Daarnaast waren 62% van de opmerkingen van zorgverleners over tijdelijke belasting/tijdsdruk positief en 29% negatief, voornamelijk vanwege de benodigde tijd voor het bewerken van AI-gegenereerde notities.

Bundy vermeldde in zijn kwalitatieve studie dat de meerderheid van de zorgverleners een vermindering van cognitieve belasting ervoer.

Should-have 2 – Administratieve werkdruk

Belangrijke uitkomsten die voor deze should-have zijn gerapporteerd, betreffen administratieve werkdruk, de tijd voor verslaglegging, de tijd besteed aan verslaglegging na werktijd en de aanvullingen in verslaglegging door de zorgverlener. Deze zijn gemeten met behulp van een Likertschaal, procentuele verschillen en beschrijvende analyses.

Administratieve werkdruk

Samenvattend: Het merendeel van de zorgprofessionals rapporteerde een verlaging van de administratieve werkdruk. De ervaren verlaging was groter bij vrouwelijke zorgprofessionals.

Twee studies rapporteerden over de administratieve werkdruk van AI-spraak applicaties (Ha, Shah_b).

In de mixed-methodsstudie van Ha gaf 68% van de zorgverleners aan dat AI-spraak applicaties de administratieve werkdruk verminderden. Het aandeel vrouwelijke zorgverleners dat een verlaging van administratieve werkdruk rapporteerde was groter dan bij mannen (67,1% vs. 53,2%).

In de kwalitatieve studie van Shah_b was 89% van de opmerkingen over de administratieve werkdruk met AI-spraak applicaties positief. Daarnaast was 91% van de opmerkingen met betrekking tot werk-privébalans positief.

Tijd voor verslaglegging

Samenvattend: De geïncludeerde wetenschappelijke studies laten een significante verlaging in de tijd voor verslaglegging zien met AI-spraak applicaties. Eén studie suggereert dat AI-spraak applicaties mogelijk verschillen in de mate van tijdsbesparing.

Zwarte literatuur

Acht studies rapporteerden over de tijd voor verslaglegging (Van Buchem, Biro_b, Shah_a, Ha, Liu, Ma, Bundy).

In de Nederlandse simulatiestudie van Van Buchem werd een significante verlaging in de mediane tijd voor verslaglegging gemeten (Seconden: 202 vs. 152; p=<0,001).

In de simulatiestudie met twee verschillende AI-spraak applicaties in diverse zorgsettings van Biro_b is een significant verschil gevonden van 81 seconden (SE 21,4) in de tijd om notities te bewerken tussen twee commerciële AI-spraak applicaties.

In de kwaliteitsverbeteringsstudie van Shah_a rapporteerde 65% van de zorgverleners een verhoogde efficiëntie bij verslaglegging. Verder ervoeren zorgverleners een mediane tijdsbesparing van 20 minuten per halve dag in de praktijk. De spreiding was echter groot, variërend van 90 minuten tijdsbesparing tot 90 minuten extra tijd voor verslaglegging.

In het simulatie onderdeel van de mixed-methodsstudie van Ha werd een significant reductie gevonden in de tijd voor verslaglegging per consult van 329 (SD 90,5) seconden naar 100 (SD 62,7) seconden. Ook was er een significante reductie te zien in de gemiddelde aantal uren voor administratieve taken per week (14,3 vs. 11,0; p=<0,05).

Een significant deel van de zorgverleners (43,5%) rapporteerde in de pre-post-vergelijkende, niet-gerandomiseerde trial van Liu een reductie in de tijd voor het verslagleggen van consulten met een AI-spraak applicatie, ten opzichte van de controlegroep (18,2%).

In de kwaliteitsverbeteringsstudie (pre-postvergelijking) van Ma is de mediane tijd voor verslaglegging significant lager na introductie van de AI-spraak applicatie (-0,57 minuten, IQR -1,3 tot -0,13). De mediane gemiddelde tijd per notitie daalde van 4,86 minuten naar 3,64 minuten na de introductie van de AI-spraak applicatie. De mediane dagelijkse verslagleggingstijd toonde een significante reductie van 6,89 minuten (IQR -22,37 tot -0,65).

In de kwalitatieve studie zonder vergelijking van Bundy rapporteerde de meerderheid van de zorgverleners een reductie in de tijd voor verslaglegging met een AI-spraak applicatie.

Grijze literatuur

In de pre-post pilot evaluatie van Haak zorgde het gebruik van Juvoly voor een significante gemiddelde afname van de consultduur (-1:54 CI (-02:56 -_-00:53)), waarbij de afname het grootste was voor de huisartsen (-02:03 minuut gemiddeld per consult). Daarnaast werd een significante gemiddelde afname van de administratie duur gemeten (-1:40 minuut; CI (-01:56 - -01:24)).

Overwerken voor verslaglegging

Samenvattend: De wetenschappelijke studies laten een reductie in tijd voor verslaglegging na werktijd zien. De reductie in tijd verschilt tussen zorgverleners, zoals verpleegkundig specialisten en huisartsen.

Zwarte literatuur

Vier primaire studies rapporteerden over de tijd in het HIS na werktijd (Ha, Liu, Ma, Bundy).

In het simulatie-onderdeel van de mixed-methodsstudie van Ha werd een significant verschil gevonden in de tijd voor verslaglegging na werktijd per week na de introductie van een AI-spraak applicatie (van 9,3 uur per week naar 6,3 uur, p=<0,05). Het percentage verpleegkundig specialisten die een reductie in verslaglegging na werktijd ervaart, is significant hoger dan bij huisartsen (84,2% vs. 56,7%).

In de pre-post vergelijkende niet-gerandomiseerde trial van Liu rapporteerde 44,7% van de zorgverleners een reductie in tijd voor verslaglegging na werktijd en 47,1% een reductie in tijd besteed in het HIS vanuit huis na de introductie van een AI-spraak applicatie. De percentage zorgverleners was significant groter dan in de controlegroep (20% en 14,5%).

In de kwaliteitsverbeteringsstudie (pre-postvergelijking) van Ma werd een significante verlaging in dagelijkse uren in het HIS na werktijd gevonden na de introductie van een AI-spraak applicatie.

In de kwalitatieve studie zonder vergelijking van Bundy rapporteerden zorgverleners een reductie in de thuis bestede tijd aan documentatie bij het gebruik van een AI-spraak applicatie, met name onder zorgverleners die eerder hun notities dicteerden.

Bijdrage van zorgverleners aan AI-genereerde verslaglegging

Samenvattend: Het bewerken van AI-gegenereerde verslagen kost minder tijd dan handmatig intypen. Zorgverleners ervoeren de hoeveelheid en de inspanning voor het bewerken van AI-gegenereerde notities als matig.

Zwarte literatuur

Twee simulatiestudies rapporteerden over de bijdrage van zorgverleners aan AI-gegenereerde verslaglegging (Biro_b, Wang).

In de studie van Biro_b werd de ervaren hoeveelheid bewerking van de notities aan twee zorgprofessionals gevraagd. Zorgverleners ervoeren dat AI-gegenereerde notities een matige hoeveelheid bewerking vereisten (Likertschaal 1-5: 2,89; SD 0,86). Verder werd de ervaren inzet van zorgprofessionals voor het aanpassen van notities uitgevraagd. Hierbij gaven zorgprofessionals aan een minimale tot gemiddelde inspanning nodig te hebben (Likertschaal 1-5: 2,73; SD 0,84). Er werd daarbij een significant verschil gevonden tussen de twee zorgprofessionals die de notities bewerkten.

Wang rapporteerde over de tijd in minuten voor het bewerken van AI-gegenereerde notities, gemeten in woorden per minuut (WPM). Het bewerken van AI-gegenereerde notities was significant sneller dan het handmatig opstellen van notities (WPM-verschil: 131,49; p=<0.0001).

Should-have 3 – Tijdigheid – Aandacht voor het gesprek zorgprofessional perspectief

Samenvattend: Op basis van de verzamelde kwalitatieve inzichten uit de wetenschappelijke studies ervaren zorgprofessionals een verbeterde communicatie met patiënten tijdens het consult met behulp van AI-spraak applicaties.

Zwarte literatuur

Drie primaire studies rapporteerden over de communicatie tussen patiënt en zorgverlener (Shah_b, Bundy, Ha).

In de kwalitatieve studie van Shah_b was 68% van de opmerkingen over patiëntbetrokkenheid positief, 21% neutraal een 11% negatief.

In de kwalitatieve studie van Bundy vond de meerderheid van de zorgverleners dat de AI-spraak applicatie zorgbezoeken met meer betrokkenheid faciliteerde.

Ook in de mixed-methodsstudie van Ha gaven zorgverleners aan dat zij zich beter konden concentreren op het gesprek en meer betrokken konden zijn tijdens het consult.

Should-have 4 – Tijdigheid – Aandacht voor het gesprek patiënt perspectief

Samenvattend: Patiënten ervaren de patiënt–zorgprofessionalrelatie over het algemeen als persoonlijker en rapporteren dat er minder tijd aan typen wordt besteed. De algemene tevredenheid van patiënten is echter vergelijkbaar met en zonder AI-spraak applicatie.

Zwarte literatuur

Twee primaire studies rapporteerden over de patiënt–zorgverlenerrelatie (Ha, Owens).

In de observationele studie van Owens werd een significant verschil gevonden ten opzichte van de huidige manier van verslaglegging: patiënten ervoeren meer aandacht van de zorgverlener (75,4%; CI 70,4%-80,3%), minder tijd besteed aan typen (78,8%; CI: 74,1-83,5%) en een persoonlijker consult (80,9%; CI: 76,4%-85,4%), gemeten met een Likertschaal. Op de Patient–Doctor Relationship Questionnaire (PDRQ-9), die communicatie en informatieoverdracht, begrip en empathie, vertrouwen en betrokkenheid en tevredenheid met de relatie meet, werd echter geen significant verschil gevonden met of zonder AI-spraak applicatie (45 vs. 45; p=0.31). Ook op de luister- en tevredenheidsschaal (score 0–100) werd geen significant verschil vastgesteld (74% vs. 68.7%; p=0,31).

In de mixed-methodsstudie van Ha ervoeren patiënten het consult met een AI-spraak applicatie eveneens als persoonlijker, met minder tijd besteed aan typen en meer mentale en fysieke aanwezigheid van de zorgverlener. Daarnaast werd vaker naar het welzijn van patiënten gevraagd. Sommige patiënten rapporteerden echter geen merkbare verandering door het gebruik van AI.

Should-have 5 – Begrijpelijkheid

Er zijn geen studies gevonden.

5 Discussie

Het onderzoek laat zien dat de meetplan indicatoren voor de huisartsenzorg (HAZ), overeenkomen met de uitkomsten gevonden in de wetenschappelijke literatuur. Dit suggereert dat deze indicatoren goed meetbaar zijn.

De meest gerapporteerde thema's in de geïncludeerde literatuur waren de kwaliteit van de samenvattingen (n=6), het werkplezier van zorgverleners (n=6), de houding van zorgverleners (n=5) en de impact op administratieve tijd (n=8). Minder vaak werd gerapporteerd over de houding van patiënten (n=2), gebruik (n=3) en patiënt-zorgverlenerrelatie vanuit het perspectief van zorgverleners (n=3) en patiënten (n=2). Voor de should-have-indicator 'Begrijpelijkheid' werden geen relevante uitkomsten gevonden.

De studies verschillen in bewijskracht op basis van het studiedesign. Twee studies hadden een design met gemiddelde bewijskracht: een niet-gerandomiseerde klinische trial en een observationele studie met een single-arm open-labelfase. De overige studies hadden een design met lage bewijskracht, zoals simulatiestudies (n=6), kwalitatieve studies (n=2), pre-post observationele studies (n=2), kwantitatieve studies (n=1) en mixed-methodsstudies (n=1). Bij de beoordeling van de methodologische kwaliteit werd duidelijk dat de helft van de studies van onvoldoende kwaliteit waren. Een deel van de simulatiestudies, de observationele en de kwantitatieve studie, waren niet representatief voor de doelgroep. Dit werd veroorzaakt door het gebruik van mockconsultaties of een beperkte zorgverlenerpopulatie, waarbij bijvoorbeeld zorgverleners met een hoge tijdsbesteding aan verslaglegging werden meegenomen. In paragraaf 1.3 'Evidence gap' werd de kwaliteit van de studies per meetplanindicator bekeken.

Daarnaast betreft de doelpopulatie in 6 van de 14 studies zowel zorgverleners uit de medisch-specialistische zorg als de huisartsenzorg. De behoeften en ervaringen van deze twee groepen zorgverleners kunnen dus verschillen, bijvoorbeeld wat betreft werkplezier en werkbelasting, evenals de subjectieve beoordeling van de kwaliteit van AI-notities, aangezien de inhoud van werk en de setting verschilt.

57% van de studies komt uit de Amerikaanse zorgsetting, waar de 'care as usual' vaak het gebruik van een menselijke schrijfhulp omvat, ook als dat niet expliciet wordt vermeld in de studie. Daarnaast is de financiële context anders in de VS: daar moeten 'billable codes' correct worden vastgelegd door de AI-spraak applicatie, wat in de Nederlandse huisartsenzorg minder relevant is. In de VS ligt de nadruk bij adoptie van AI-spraak applicaties daarom op 'billing compliance' en efficiëntie, terwijl in Nederland de focus meer ligt op reductie van burn-out en efficiëntie.

Verder heeft de Nederlandse huisartsenzorg een breed medisch generalistisch karakter, breder van in de VS, waardoor er behoefte is aan AI-spraak applicaties met brede medische terminologie.

Zes studies gaven geen inzicht in welke applicatie werd gebruikt, terwijl zeven studies gebruikmaakten van DAX, dat in Nederland (nog) niet wordt toegepast. Alleen de Nederlandse studies maakte gebruik van een applicatie ontwikkeld voor de Nederlandse setting (Autoscribe en Juvoly). AI-spraak applicaties specifiek voor de Nederlandse of Europese zorgsetting zijn dus nog nauwelijks onderzocht of gerapporteerd in de wetenschappelijke literatuur.

Daarnaast bestond de helft van de geïncludeerde studies uit simulaties. Hierdoor is het moeilijk om conclusies te trekken over de impact van AI-spraak applicaties in de dagelijkse praktijk.

6 Conclusie

Naar de meeste meetplan indicatoren is onderzoek gedaan, echter de kwaliteit was regelmatig beperkt of niet representatief voor de Nederlandse huisartsenzorg. In de Nederlandse huisartsenzorg is de behoefte aan AI-verslaglegging anders dan in andere landen, juist vanwege het medisch generalistische karakter van deze zorg. Daarom is het moeilijk om resultaten uit buitenlandse studies direct over te nemen voor de Nederlandse situatie. Hoe geschikt een AI-notitie is voor een bepaalde setting bepaalt immers ook de kwaliteit van de samenvattingen. En dat bepaalt weer hoeveel tijd zorgprofessionals nodig hebben om notities aan te passen. Het heeft ook invloed op hun werkplezier en op hoe patiënten en zorgprofessionals denken over het gebruik van AI-spraaktoepassingen.

In Bijlage F bevindt zich een overzicht van toepassingen die in de Nederlandse HAZ in ontwikkeling zijn, met daarbij de claims die zij doen over hun inzet op administratieve werkdruk, werkplezier en de kwaliteit van de toepassingen. Deze claims konden echter niet verder geëvalueerd worden vanwege het gebrek aan gegevens.

7 Referenties

Anderson, T. N., Mohan, V., Dorr, D. A., Ratwani, R. M., Biro, J. M., & Gold, J. A. (2025). Evaluating the Quality and Safety of Ambient Digital Scribe Platforms Using Simulated Ambulatory Encounters. https://doi.org/10.2139/ssrn.5255300

Biro, J., Handley, J. L., Cobb, N. K., Kottamasu, V., Collins, J., Krevat, S., & Ratwani, R. M. (2025). Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study. Journal of Medical Internet Research, 27, e64993. https://doi.org/10.2196/64993

Biro, J., Handley, J. L., Mickler, J., Reddy, S., Kottamasu, V., Ratwani, R. M., & Cobb, N. K. (2025). The value of simulation testing for the evaluation of ambient digital scribes: A case report. Journal of the American Medical Informatics Association, 32(5), 928–931. https://doi.org/10.1093/jamia/ocaf052

Bundy, H., Gerhart, J., Baek, S., Connor, C. D., Isreal, M., Dharod, A., Stephens, C., Liu, T.-L., Hetherington, T., & Cleveland, J. (2024). Can the Administrative Loads of Physicians be Alleviated by AI-Facilitated Clinical Documentation? Journal of General Internal Medicine, 39(15), 2995–3000. https://doi.org/10.1007/s11606-024-08870-z

Ha, E. (2024, July 31). CLINICAL EVALUATION OF ARTIFICIAL INTELLIGENCE AND AUTOMATION TECHNOLOGY TO REDUCE ADMINISTRATIVE BURDEN IN PRIMARY CARE. Centre for Digital Health Evaluation (CDHE) at the Women's College Hospital Institute for Health System Solutions and Virtual Care (WIHV). https://km4s.ca/wp-content/uploads/Clinical-evaluation-of-artificial-intelligence-and-automation-technology-to-reduce-administrative-burden-in-primary-care-2024.pdf

Haak, M., Zuidwijk, Y., de Jong-Gerding, E., van der Worn, H., Blanker, M. H., & Loohuis, A. M. M. (2025, June). Effect van AI-consultregistratie (Juvoly) op efficientie, kwaliteit en werkplezier in de spreekkamer. https://www.dokterdrenthe.nl/assets/uploads/Documenten/Volledig-Rapport_Effect-AI-spraakherkenning-in-de-spreekkamer_juni2025.pdf

Hose, B.-Z., Handley, J. L., Biro, J., Reddy, S., Krevat, S., Hettinger, A. Z., & Ratwani, R. M. (2025). Development of a Preliminary Patient Safety Classification System for Generative AI. BMJ Quality & Safety, 34(2), 130–132. https://doi.org/10.1136/bmjqs-2024-017918

Liu, T.-L., Hetherington, T. C., Stephens, C., McWilliams, A., Dharod, A., Carroll, T., & Cleveland, J. A. (2024). AI-Powered Clinical Documentation and Clinicians' Electronic Health Record Experience: A Nonrandomized Clinical Trial. JAMA Network Open, 7(9), e2432460. https://doi.org/10.1001/jamanetworkopen.2024.32460

Ma, S. P., Liang, A. S., Shah, S. J., Smith, M., Jeong, Y., Devon-Sand, A., Crowell, T., Delahaie, C., Hsia, C., Lin, S., Shanafelt, T., Pfeffer, M. A., Sharp, C., & Garcia, P. (2025). Ambient artificial intelligence scribes: Utilization and impact on documentation time. Journal of the American Medical Informatics Association, 32(2), 381–385. https://doi.org/10.1093/jamia/ocae304

Misurac, J., Knake, L. A., & Blum, J. M. (2025). The Effect of Ambient Artificial Intelligence Notes on Provider Burnout. Applied Clinical Informatics, 16(02), 252–258. https://doi.org/10.1055/a-2461-4576

Owens, L. M., Wilda, J. J., Grifka, R., Westendorp, J., & Fletcher, J. J. (2024). Effect of Ambient Voice Technology, Natural Language Processing, and Artificial Intelligence on the Patient–Physician Relationship. Applied Clinical Informatics, 15(04), 660–667. https://doi.org/10.1055/a-2337-4739

Shah, S. J., Crowell, T., Jeong, Y., Devon-Sand, A., Smith, M., Yang, B., Ma, S. P., Liang, A. S., Delahaie, C., Hsia, C., Shanafelt, T., Pfeffer, M. A., Sharp, C., Lin, S., & Garcia, P. (2025). Physician Perspectives on Ambient AI Scribes. JAMA Network Open, 8(3), e251904. https://doi.org/10.1001/jamanetworkopen.2025.1904

Shah, S. J., Devon-Sand, A., Ma, S. P., Jeong, Y., Crowell, T., Smith, M., Liang, A. S., Delahaie, C., Hsia, C., Shanafelt, T., Pfeffer, M. A., Sharp, C., Lin, S., & Garcia, P. (2025). Ambient artificial intelligence scribes: Physician burnout and perspectives on usability and documentation burden. Journal of the American Medical Informatics Association, 32(2), 375–380. https://doi.org/10.1093/jamia/ocae295

Van Buchem, M. M., Kant, I. M. J., King, L., Kazmaier, J., Steyerberg, E. W., & Bauer, M. P. (2024). Impact of a Digital Scribe System on Clinical Documentation Time and Quality: Usability Study. JMIR AI, 3, e60020. https://doi.org/10.2196/60020

Wang, J., Lavender, M., Hoque, E., Brophy, P., & Kautz, H. (2021). A patient-centered digital scribe for automatic medical documentation. JAMIA Open, 4(1), ooab003. https://doi.org/10.1093/jamiaopen/ooab003

Bijlagen

Bijlage A – Meetplan

Hoofddomein	Omschrijving algehele indicator	Specificatie van claim	Acceptatiegrens
Must have – Veiligheid	De mate waarin SOLL veilig is, zowel op gebied van zorg en gezondheid.	De SOEP- weergave van het consult middels SGR is een bruikbare* weergave van het gesprek in de spreekkamer.	Minimaal gelijk aan IST
Must have - Acceptatie	Hoe de beoogde individuele eindgebruikers reageren op SOLL (in vergelijking met IST): ervaringen/percepties (bijv. nut, gemak, vertrouwen, houding, intentie, aanraden van het proces aan anderen).	Patiënten staan niet negatief tegenover de inzet van SGR door zorgprofessionals.	Minimaal gelijk aan IST
Should have – Adoptie (zorgprofessionals /onder-steuners)	Hoe verschillende (representatieve) betrokkenen bij de implementatie (bijv. zorgprofessionals en mantelzorgers) reageren op SOLL (in vergelijking met IST) in gedrag, perceptie, of beide.	Zorgprofessionals staan positief tegenover het gebruik van SGR.	Minimaal gelijk aan IST
Should have – Werkplezier, werktevredenheid, of werkdruk (focus werkplezier)	Hoe de zorgprofessional de arbeidsomstandigheden ervaart.	Door gebruik van SGR ervaart de zorgprofessional meer werkplezier.	Minimaal gelijkaan IST
Should have - Werkplezier, werktevredenheid, of werkdruk (focus werkdruk)	Hoe de zorgprofessional de arbeidsomstandigheden ervaart	Inzet van SGR vermindert administratieve druk voor de zorgprofessional tijdens het consult.	Minimaal gelijkaan IST
Should have - Tijdigheid	De mate waarin eindgebruikers (van SOLL) de juiste zorg of ondersteuning op het juiste moment door de juiste professional en op de juiste plek ontvangen.	Door inzet van SGR hebben zorgprofessional en patiënt meer aandacht voor het gesprek. De **zorgprofessionals** ervaren door SGR een verbetering van de kwaliteit van het gesprek.**	Minimaal gelijkaan IST
Should have - Tijdigheid	De mate waarin eindgebruikers (van SOLL) de juiste zorg of ondersteuning op het juiste moment door de juiste professional en op de juiste plek ontvangen.	Door inzet van SGR hebben zorgprofessional en patiënten meer aandacht voor het gesprek. De **patiënten** zijn tevreden over de kwaliteit van het gesprek **	Minimaal gelijkaan IST
Should have -Toegankelijkheid	De mate waarin het proces aangepast is aan individuele verschillen tussen eindgebruikers. Hierbij kunnen kennis/vaardig-heden voor gebruik (bijv: visueel, auditief, taal) en voorkeuren/overtuigingen (bijv. geloof, identiteit) van belang zijn.	De leesbaarheid en begrijpelijkheid van de rapportage is gelijk of beter dan voorheen.	Minimaal gelijkaan IST

Bijlage B – Criteria grijze literatuur

Grijze literatuur – Criteria voor inclusie/exclusie

Selectiecriteria (inclusie)	
Het document bevat resultaten/ uitkomsten van een (pilot/implementatie/onderzoeks) traject*	Ja/Nee
Het document bevat een methode**	Ja/Nee
Het document beschrijf de juiste setting (sector en SOLL)	Ja/Nee
Het document heeft een auteur / afzender***	Ja/Nee
Het document is recent (< 5 jaar)	Ja/Nee
Inclusie bij 5x JA	
* mag ook kwalitatieve data beschrijven vanuit interviews of focusgroepen	
** het is helder waar de data vandaan komt en wat ermee gedaan is	
*** kan ook een instelling, bedrijf of leverancier zijn	
Exclusie	
Documenten voor promotiedoeleinden zonder data/onderbouwing	
Documenten zonder uitkomsten/resultaten	
Documenten/presentaties met uitkomsten/resultaten die niet herleidbaar zijn tot het brondocument	
Handleidingen bij toepassingen	
Documenten die de productontwikkeling/verantwoording beschrijven	
Studieprotocollen	
Achtergrondinformatie (bijvoorbeeld bij werkingsmechanisme techniek SGR)	
Hypothetische business case beschrijvingen	

Bijlage C – Zoekstrategie

Tabel 1 - OVID/Medline 11 jul 2025

No.	Query	Results
#1	exp "Speech Recognition Software"/ or (google-speech or ambient-scribe* or digital-scribe* or AI-spraakapplicatie* or virtual-scribe* or deepscribe* or deep-scribe* or ai-generated-scribe* or machine-learning-scribe* or automatic-scribe* or artificial-intelligence-scribe* or automatic-transcri* or ai-transcri* or virtual-transcri* or deeptranscri* or deep-transcri* or ambient-transcri* or machine-learning-transcri* or artificial-intelligence-transcri* or ai-generated-transcri* or sassi or vocera or ((speech or command* or voice-recogn*) adj3 (interface* or software* or system* or automat* or smartwatch or smart-watch or technolog*)) or ((dictat* or speech) adj3 (text or command*) adj3 (executi* or functional* or interface*)) or voice-activated-system* or "attendi" or "tell james" or "medendo assistent" or "juvoly quickconsult" or "digitale assistent" or "autoscriber" or "health talk" or "ontzorgd" or "verticai" or "(nuance adj3 microsoft)" or "tonos" or "cato" or "elanza" or "smart planning" or "youplan" or "voca zorg" or "epic assistent" or "robin scribe" or "consult assistent" or "g2speech" or "ourmind" or "wellcom health").ti,ab,kf.	**6496**
#2	"Electronic Health Records"/ or "Medical Records"/ or "Medical Records Systems, Computerized"/ or "Documentation"/ or (medical-record* or electronic-medical-record* or electronic-health-record* or computerized-medical-record* or computerised-medical-record* or computerized-patient-medical-records* or computerised-patient-medical-records* or electronic-health-record* or automated-medical-record* or electronic-patient-record* or electronic-client-record* or electronic-client-documentation* or electronic-patient-documentation* or clinical-documentation* or physician-documentation* or pdqi* or medical-note* or nurse-documentation* or medical-information-management* or automated-report*).ti,ab,kf.	**306443**
#3	1 and 2	**399**

Tabel 2 - Embase.com 13 jul 2025

#	Searches	Results
#4	#3 NOT ('Conference Abstract'/it OR 'Conference Paper'/it)	**415**
#3	#1 AND #2	**478**
#2	'electronic health record'/de OR 'electronic medical record'/de OR 'electronic patient record'/de OR 'medical record'/de OR 'electronic medical record system'/de OR 'medical documentation'/de OR 'medicalrecord*':ti,ab,kw OR 'electronic medical record*':ti,ab,kw OR 'computerized medical record*':ti,ab,kw OR 'computerised medical record*':ti,ab,kw OR 'computerized patient medical records*':ti,ab,kw OR 'computerised patient medical records*':ti,ab,kw OR 'electronic health record*':ti,ab,kw OR 'automatedmedical record*':ti,ab,kw OR 'electronic patient record*':ti,ab,kw OR 'electronic client record*':ti,ab,kw OR 'electronic client documentation*':ti,ab,kw OR 'electronic patient documentation*':ti,ab,kw OR 'clinicaldocumentation*':ti,ab,kw OR 'physician documentation*':ti,ab,kw OR pdqi*:ti,ab,kw OR 'medicalnote*':ti,ab,kw OR 'nurse documentation*':ti,ab,kw OR 'medical information management*':ti,ab,kw OR 'automated report*':ti,ab,kw	**561666**

#1	'automatic speech recognition'/exp OR 'google speech':ti,ab,kw OR 'ambient scribe*':ti,ab,kw OR 'digital scribe*':ti,ab,kw OR 'ai scribe*':ti,ab,kw OR 'virtual scribe*':ti,ab,kw OR deepscribe*:ti,ab,kw OR 'deepscribe*':ti,ab,kw OR 'ai generated scribe*':ti,ab,kw OR 'machine learning scribe*':ti,ab,kw OR 'automatic scribe*':ti,ab,kw OR 'artificial intelligence scribe*':ti,ab,kw OR 'automatic transcri*':ti,ab,kw OR 'ai transcri*':ti,ab,kw OR 'virtual transcri*':ti,ab,kw OR deeptranscri*:ti,ab,kw OR 'deep transcri*':ti,ab,kw OR 'ambient transcri*':ti,ab,kw OR 'machine learning transcri*':ti,ab,kw OR 'artificial intelligence transcri*':ti,ab,kw OR 'ai generated transcri*':ti,ab,kw OR sassi:ti,ab,kw OR vocera:ti,ab,kw OR (((speech OR command* OR 'voice recogn*') NEAR/3 (interface* OR software* OR system* OR automat* OR smartwatch OR 'smart watch' OR technolog*)):ti,ab,kw) OR (((dictat* OR speech) NEAR/3 (text OR command*) NEAR/3 (executi* OR functional* OR interface*)):ti,ab,kw) OR 'voice activated system*':ti,ab,kw OR 'attendi':ti,ab,kwOR 'tell james':ti,ab,kw OR 'medendo assistent':ti,ab,kw OR 'juvoly quickconsult':ti,ab,kw OR 'digitale assistent':ti,ab,kw OR 'autoscriber':ti,ab,kw OR 'health talk':ti,ab,kw OR 'ontzorgd':ti,ab,kw OR 'verticai':ti,ab,kw OR ('nuance' NEAR/3 'microsoft'):ti,ab,kw OR 'tonos':ti,ab,kw OR 'cato':ti,ab,kw OR 'elanza':ti,ab,kw OR 'smart planning':ti,ab,kw OR 'youplan':ti,ab,kw OR 'voca zorg':ti,ab,kw OR 'epicassistent':ti,ab,kw OR 'robin scribe':ti,ab,kw OR 'consult assistent':ti,ab,kw OR 'g2speech':ti,ab,kw OR 'ourmind':ti,ab,kw OR 'wellcom health':ti,ab,kw	**8111**

Tabel 3 - Elsevier/Scopus 13 jul 2025

#	Searches	Results
#3	#1 AND #2	**258**
#2	TITLE-ABS (medical-record* OR electronic-medical-record* OR electronic-health-record* OR computerized-medical-record* OR computerised-medical-record* OR computerized-patient-medical-records* OR computerised-patient-medical-records* OR electronic-health-record* OR automated-medical-record* OR electronic-patient-record* OR electronic-client-record* OR electronic-client-documentation* OR electronic-patient-documentation* OR clinical-documentation* OR physician-documentation* OR pdqi* OR medical-note* OR nurse-documentation* OR medical-information-management* OR automated-report*) OR AUTHKEY (medical-record* OR electronic-medical-record* OR electronic-health-record* OR computerized-medical-record* OR computerised-medical-record* OR computerized-patient-medical-records* OR computerised-patient-medical-records* OR electronic-health-record* OR automated-medical-record* OR electronic-patient-record* OR electronic-client-record* OR electronic-client-documentation* OR electronic-patient-documentation* OR clinical-documentation* OR physician-documentation* OR pdqi* OR medical-note* OR nurse-documentation* OR medical-information-management* OR automated-report*)	**249600**
#1	TITLE-ABS(google-speech OR ambient-scribe* OR digital-scribe* OR AI-spraak applicatie* OR virtual-scribe* OR deepscribe* OR deep-scribe* OR ai-generated-scribe* OR machine-learning-scribe* OR automatic-scribe* OR artificial-intelligence-scribe* OR automatic-transcri* OR ai-transcri* OR virtual-transcri* OR deeptranscri* OR deep-transcri* OR ambient-transcri* OR machine-learning-transcri* OR artificial-intelligence-transcri* OR ai-generated-transcri* OR sassi OR vocera OR ((speech OR command* OR voice-recogn*) W/3 (interface* OR software* OR system* OR automat* OR smartwatch OR smart-watch OR technolog*)) OR ((dictat* OR speech) W/3 (text OR command*) W/3 (executi* OR functional* OR interface*)) OR voice-activated-system* OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" W/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robin scribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health") OR AUTHKEY(google-speech OR ambient-scribe* OR digital-scribe* OR AI-spraak applicatie* OR virtual-scribe* OR deepscribe* OR deep-scribe* OR ai-generated-scribe* OR machine-learning-scribe* OR automatic-scribe* OR artificial-intelligence-scribe* OR automatic-transcri* OR ai-transcri* OR virtual-transcri* OR deeptranscri* OR deep-transcri* OR ambient-transcri* OR machine-learning-transcri* OR artificial-intelligence-transcri* OR ai-generated-transcri* OR sassi OR vocera OR ((speech OR command* OR voice-recogn*) W/3 (interface* OR software* OR system* OR automat* OR smartwatch OR smart-watch OR technolog*)) OR ((dictat* OR speech) W/3 (text OR command*) W/3 (executi* OR functional* OR interface*)) OR voice-activated-system* OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" W/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robin scribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health")	**77119**

Tabel 4 - Clarivate Analytics/Web of Science Core Collection (13 juli 2025)

#	Searches	Results
#3	#2 AND #1	**139**
#2	TS=("medical record*" OR "electronic medical record*" OR "electronic health record*" OR "computerizedmedical record*" OR "computerised medical record*" OR "computerized patient medical records*" OR "computerised patient medical records*" OR "electronic health record*" OR "automated medical record*" OR "electronic patient record*" OR "electronic client record*" OR "electronic client documentation*" OR "electronic patient documentation*" OR "clinical documentation*" OR "physician documentation*" OR "pdqi*" OR "medical note*" OR "nurse documentation*" OR "medical information management*" OR "automated report*")	**203319**
#1	TS=("google speech" OR "ambient scribe*" OR "digital scribe*" OR "ai scribe*" OR "virtual scribe*" OR "deepscribe*" OR "deep scribe*" OR "ai generated scribe*" OR "machine learning scribe*" OR "automatic scribe*" OR "artificial intelligence scribe*" OR "automatic transcri*" OR "ai transcri*" OR "virtual transcri*" OR "deeptranscri*" OR "deep transcri*" OR "ambient transcri*" OR "machine learning transcri*" OR "artificial intelligence transcri*" OR "ai generated transcri*" OR "sassi" OR "vocera" OR (("speech" OR "command*" OR "voice recogn*") NEAR/3 ("interface*" OR "software*" OR "system*" OR "automat*" OR "smartwatch" OR "smart watch" OR "technolog*")) OR (("dictat*" OR "speech") NEAR/3 ("text" OR "command*") NEAR/3 ("executi*" OR "functional*" OR "interface*")) OR "voice activated system*" OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" NEAR/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robinscribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health")	**21445**

Bijlage bij zoekstrategieën Spraakgestuurd rapporteren + update (13 juli 2025)

OVID/Medline 13 jul 2025

Ovid MEDLINE(R) ALL <1946 to July 11, 2025>

1	exp "Speech Recognition Software"/ or (google-speech or ambient-scribe* or digital-scribe* or AI-spraak applicatie* or virtual-scribe* or deepscribe* or deep-scribe* or ai-generated-scribe* or machine-learning-scribe* or automatic-scribe* or artificial-intelligence-scribe* or automatic-transcri* or ai-transcri* or virtual-transcri* or deeptranscri* or deep-transcri* or ambient-transcri* or machine-learning-transcri* or artificial-intelligence-transcri* or ai-generated-transcri* or sassi or vocera or ((speech or command* or voice-recogn*) adj3 (interface* or software* or system* or automat* or smartwatch or smart-watch or technolog*)) or ((dictat* or speech) adj3 (text or command*) adj3 (executi* or functional* or interface*)) or voice-activated-system* or "attendi" or "tell james" or "medendo assistent" or "juvoly quickconsult" or "digitale assistent" or "autoscriber" or "health talk" or "ontzorgd" or "verticai" or "(nuance adj3 microsoft)" or "tonos" or "cato" or "elanza" or "smart planning" or "youplan" or "voca zorg" or "epic assistent" or "robin scribe" or "consult assistent" or "g2speech" or "ourmind" or "wellcom health").ti,ab,kf.	**6496**
2	"Electronic Health Records"/ or "Medical Records"/ or "Medical Records Systems, Computerized"/ or "Documentation"/ or (medical-record* or electronic-medical-record* or electronic-health-record* or computerized-medical-record* or computerised-medical-record* or computerized-patient-medical-records* or computerised-patient-medical-records* or electronic-health-record* or automated-medical-record* or electronic-patient-record* or electronic-client-record* or electronic-client-documentation* or electronic-patient-documentation* or clinical-documentation* or physician-documentation* or pdqi* or medical-note* or nurse-documentation* or medical-information-management* or automated-report*).ti,ab,kf.	**306443**
3	1 and 2	**399**

Embase.com 13 jul 2025

No.	Query	Results
#4	#3 NOT ('Conference Abstract'/it OR 'Conference Paper'/it)	**415**
#3	#1 AND #2	**478**
#2	'electronic health record'/de OR 'electronic medical record'/de OR 'electronic patient record'/de OR 'medical record'/de OR 'electronic medical record system'/de OR 'medical documentation'/de OR 'medical record*':ti,ab,kw OR 'electronic medical record*':ti,ab,kw OR 'computerized medical record*':ti,ab,kw OR 'computerised medical record*':ti,ab,kw OR 'computerized patient medical records*':ti,ab,kw OR 'computerised patient medical records*':ti,ab,kw OR 'electronic health record*':ti,ab,kw OR 'automated medical record*':ti,ab,kw OR 'electronic patient record*':ti,ab,kw OR 'electronic client record*':ti,ab,kw OR 'electronic client documentation*':ti,ab,kw OR 'electronic patient documentation*':ti,ab,kw OR 'clinical documentation*':ti,ab,kw OR 'physician documentation*':ti,ab,kw OR pdqi*:ti,ab,kw OR 'medical note*':ti,ab,kw OR 'nurse documentation*':ti,ab,kw OR 'medical information management*':ti,ab,kw OR 'automated report*':ti,ab,kw	**561666**
#1	'automatic speech recognition'/exp OR 'google speech':ti,ab,kw OR 'ambient scribe*':ti,ab,kw OR 'digital scribe*':ti,ab,kw OR 'ai scribe*':ti,ab,kw OR 'virtual scribe*':ti,ab,kw OR deepscribe*:ti,ab,kw OR 'deep scribe*':ti,ab,kw OR 'ai generated scribe*':ti,ab,kw OR 'machine learning scribe*':ti,ab,kw OR 'automatic scribe*':ti,ab,kw OR 'artificial intelligence scribe*':ti,ab,kw OR 'automatic transcri*':ti,ab,kw OR 'ai transcri*':ti,ab,kw OR 'virtual transcri*':ti,ab,kw OR deeptranscri*:ti,ab,kw OR 'deep transcri*':ti,ab,kw OR 'ambient transcri*':ti,ab,kw OR 'machine learning transcri*':ti,ab,kw OR 'artificial intelligence transcri*':ti,ab,kw OR 'ai generated transcri*':ti,ab,kw OR sassi:ti,ab,kw OR vocera:ti,ab,kw OR (((speech OR command* OR 'voice recogn*') NEAR/3 (interface* OR software* OR system* OR automat* OR smartwatch OR 'smart watch' OR technolog*)):ti,ab,kw) OR (((dictat* OR speech) NEAR/3 (text OR command*) NEAR/3 (executi* OR functional* OR interface*)):ti,ab,kw) OR 'voice activated system*':ti,ab,kw OR 'attendi':ti,ab,kw OR 'tell james':ti,ab,kw OR 'medendo assistent':ti,ab,kw OR 'juvoly quickconsult':ti,ab,kw OR 'digitale assistent':ti,ab,kw OR 'autoscriber':ti,ab,kw OR 'health talk':ti,ab,kw OR 'ontzorgd':ti,ab,kw OR 'verticai':ti,ab,kw OR ('nuance' NEAR/3 'microsoft'):ti,ab,kw OR 'tonos':ti,ab,kw OR 'cato':ti,ab,kw OR 'elanza':ti,ab,kw OR 'smart planning':ti,ab,kw OR 'youplan':ti,ab,kw OR 'voca zorg':ti,ab,kw OR 'epic assistent':ti,ab,kw OR 'robin scribe':ti,ab,kw OR 'consult assistent':ti,ab,kw OR 'g2speech':ti,ab,kw OR 'ourmind':ti,ab,kw OR 'wellcom health':ti,ab,kw	**8111**

Elsevier/Scopus 13 jul 2025

3	**#1 AND #2**	**258**
2	TITLE-ABS (medical-record* OR electronic-medical-record* OR electronic-health-record* OR computerized-medical-record* OR computerised-medical-record* OR computerized-patient-medical-records* OR computerised-patient-medical-records* OR electronic-health-record* OR automated-medical-record* OR electronic-patient-record* OR electronic-client-record* OR electronic-client-documentation* OR electronic-patient-documentation* OR clinical-documentation* OR physician-documentation* OR pdqi* OR medical-note* OR nurse-documentation* OR medical-information-management* OR automated-report*) OR AUTHKEY (medical-record* OR electronic-medical-record* OR electronic-health-record* OR computerized-medical-record* OR computerised-medical-record* OR computerized-patient-medical-records* OR computerised-patient-medical-records* OR electronic-health-record* OR automated-medical-record* OR electronic-patient-record* OR electronic-client-record* OR electronic-client-documentation* OR electronic-patient-documentation* OR clinical-documentation* OR physician-documentation* OR pdqi* OR medical-note* OR nurse-documentation* OR medical-information-management* OR automated-report*)	**249600**
1	TITLE-ABS(google-speech OR ambient-scribe* OR digital-scribe* OR AI-spraak applicatie* OR virtual-scribe* OR deepscribe* OR deep-scribe* OR ai-generated-scribe* OR machine-learning-scribe* OR automatic-scribe* OR artificial-intelligence-scribe* OR automatic-transcri* OR ai-transcri* OR virtual-transcri* OR deeptranscri* OR deep-transcri* OR ambient-transcri* OR machine-learning-transcri* OR artificial-intelligence-transcri* OR ai-generated-transcri* OR sassi OR vocera OR ((speech OR command* OR voice-recogn*) W/3 (interface* OR software* OR system* OR automat* OR smartwatch OR smart-watch OR technolog*)) OR ((dictat* OR speech) W/3 (text OR command*) W/3 (executi* OR functional* OR interface*)) OR voice-activated-system* OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" W/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robin scribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health") OR AUTHKEY(google-speech OR ambient-scribe* OR digital-scribe* OR AI-spraak applicatie* OR virtual-scribe* OR deepscribe* OR deep-scribe* OR ai-generated-scribe* OR machine-learning-scribe* OR automatic-scribe* OR artificial-intelligence-scribe* OR automatic-transcri* OR ai-transcri* OR virtual-transcri* OR deeptranscri* OR deep-transcri* OR ambient-transcri* OR machine-learning-transcri* OR artificial-intelligence-transcri* OR ai-generated-transcri* OR sassi OR vocera OR ((speech OR command* OR voice-recogn*) W/3 (interface* OR software* OR system* OR automat* OR smartwatch OR smart-watch OR technolog*)) OR ((dictat* OR speech) W/3 (text OR command*) W/3 (executi* OR functional* OR interface*)) OR voice-activated-system* OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" W/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robin scribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health")	**77119**

Clarivate Analytics/Web of Science Core Collection 13 jul 2025

3	**#2 AND #1**	**139**
2	TS=("medical record*" OR "electronic medical record*" OR "electronic health record*" OR "computerized medical record*" OR "computerised medical record*" OR "computerized patient medical records*" OR "computerised patient medical records*" OR "electronic health record*" OR "automated medical record*" OR "electronic patient record*" OR "electronic client record*" OR "electronic client documentation*" OR "electronic patient documentation*" OR "clinical documentation*" OR "physician documentation*" OR "pdqi*" OR "medical note*" OR "nurse documentation*" OR "medical information management*" OR "automated report*")	**203319**
1	TS=("google speech" OR "ambient scribe*" OR "digital scribe*" OR "ai scribe*" OR "virtual scribe*" OR "deepscribe*" OR "deep scribe*" OR "ai generated scribe*" OR "machine learning scribe*" OR "automatic scribe*" OR "artificial intelligence scribe*" OR "automatic transcri*" OR "ai transcri*" OR "virtual transcri*" OR "deeptranscri*" OR "deep transcri*" OR "ambient transcri*" OR "machine learning transcri*" OR "artificial intelligence transcri*" OR "ai generated transcri*" OR "sassi" OR "vocera" OR (("speech" OR "command*" OR "voice recogn*") NEAR/3 ("interface*" OR "software*" OR "system*" OR "automat*" OR "smartwatch" OR "smart watch" OR "technolog*")) OR (("dictat*" OR "speech") NEAR/3 ("text" OR "command*") NEAR/3 ("executi*" OR "functional*" OR "interface*")) OR "voice activated system*" OR "attendi" OR "tell james" OR "medendo assistent" OR "juvoly quickconsult" OR "digitale assistent" OR "autoscriber" OR "health talk" OR "ontzorgd" OR "verticai" OR ("nuance" NEAR/3 "microsoft") OR "tonos" OR "cato" OR "elanza" OR "smart planning" OR "youplan" OR "voca zorg" OR "epic assistent" OR "robin scribe" OR "consult assistent" OR "g2speech" OR "ourmind" OR "wellcom health")	**21445**

Zoekresultaten overzicht

Database	Aantal treffers vóór ontdubbelen (2024)	Aantal treffers ná ontdubbelen (2024)	Aantal treffers vóór ontdubbelen (2025)	Aantal treffers ná ontdubbelen (2025)
Medline	16 SR + 343 niet-SR	16 SR + 341 niet-SR	399	35
Embase	362	107	415	32
Scopus	219	97	258	17
WoS	115	8	139	2
Totaal	1055	569	1211	86

Bijlage D – Tabel studies per meetplanindicator

		Must-haves			Should-haves					
ID	Eerste auteur	1 Kwaliteit	2 Houding patiënten	3 Houding zorg-verlener	1 Gebruik	2 Werk-plezier	3 Admini-stratieve tijd	4 Patient-zorg-verlener-relatie (zorg-verlener)	5 Patient-zorg-verlener-relatie (patient)	6 Begrijpelijk-heid
1	van Buchem	1	0	1	0	0	1	0	0	0
2	Biro_a	1	0	0	0	0	0	0	0	0
3	Biro_b	0	0	0	0	0	1	0	0	0
4	Shah_a	1	0	1	1	1	1	0	0	0
5	Shah_b	1	1	1	0	1	0	1	0	0
6	Anderson	0	0	0	0	0	0	0	0	0
7	Ha	1	1	1	1	1	1	1	1	0
8	Hose	0	0	0	0	0	0	0	0	0
9	Owens	0	0	0	0	0	0	0	1	0
10	Wang	0	0	0	0	0	1	0	0	0
11	Liu	0	0	0	0	1	1	0	0	0
12	Ma	0	0	0	1	0	1	0	0	0
13	Bundy	1	0	1	0	1	1	1	0	0
14	Misurac	0	0	0	0	1	0	0	0	0

Bijlage E – Evidence gap

Definitie

Wat is een 'evidence gap' en hoe is deze in dit onderzoek bepaald?

Er is óf helemaal geen bewijs, óf het bestaande bewijs is onvoldoende, van lage kwaliteit, of niet relevant voor de onderzoeksvraag.

De 'evidence gap' in de huidige literatuur werd bepaald op basis van de bewijskracht van de studiedesigns, de methodologische kwaliteit van de studies, het al dan niet aanwezig zijn van vergelijkingen en de steekproefgrootte. Als er voldoende bewijs beschikbaar is, is aanvullend onderzoek niet nodig om de indicatoren te beoordelen op basis van een vooraf vastgestelde acceptatiegrens.

Uiteindelijk wordt de 'evidence gap' gezamenlijk met het veld vastgesteld. Dit overzicht dient ter informatie.

Overzicht

Voor alle relevante indicatoren voor SGR in de HAZ werd wetenschappelijke literatuur gevonden, behalve voor de should-have-indicator 'Begrijpelijkheid'.

De **kwaliteit van de AI-gegenereerde samenvatting** werd beoordeeld op algemene kwaliteit en op specifieke aspecten: accuraatheid, relevantie, organisatie en correctheid van de samenvatting (inclusief bias, fouten en hallucinaties) in acht studies; vier daarvan hadden een voldoende methodologische kwaliteit (zie tabel). Van de studies met voldoende methodologische kwaliteit werd maar in een studies de kwaliteit van AI-spraak applicaties vergeleken met reguliere verslaglegging. Deze studie had een design met lagere bewijskracht (mixed-methodsstudie) en een grote steekproef (n=155).

De studies onderzochten echter vooral de subjectieve kwaliteit van de door AI gegenereerde samenvattingen, bepaald door zorgverleners via handmatige beoordeling, bijvoorbeeld met instrumenten zoals de PDQI-9, een Likert-schaal of interviews. Maar de onderliggende LLM's zijn niet-deterministisch (de output kan bij dezelfde input telkens verschillen) en zijn onderhevig aan frequente versiewijzigingen, die de resultaten in de tijd sterk kunnen beïnvloeden. Hierdoor is het onrealistisch om uitsluitend te vertrouwen op continue feedback van eindgebruikers als waarborg voor kwaliteit. Sterker nog, dit kan de werkdruk van zorgprofessionals juist verhogen. De validatie van AI-spraak applicaties verschuift hiermee dus grotendeels naar de implementatiefase. Er zijn nog onvoldoende studies bekend waarin LLM's zelf (gedeeltelijk) de kwaliteitscontrole doet, door automatisch de kwaliteit van gegenereerde samenvattingen te toetsen. In Nederland wordt om die reden door het RIGH:T consortium onderzoek gedaan naar een geschikt validatieframework om de kwaliteit en veiligheid van verschillende AI-spraak applicaties te kunnen beoordelen. Bron: https://bio.link/rightconsortium.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Bundy	Kwalitatieve studie	Nee	12	Voldoende
Biro _a	Simulatiestudie	Nee	11	Voldoende
Shah_a	Kwantitatieve studie	Ja	48	Onvoldoende
Shah_b	Kwalitatieve studie	Nee	22	Voldoende
Ha	Mixed-methodsstudie	Ja (kwalitatief)	155 zorgaanbieders (enquête), 11 zorgaanbieders (interviews)	Voldoende
Van Buchem	Simulatiestudie	Ja	27	Onvoldoende
Hose	Simulatiestudie	Nee	11	Onvoldoende
Anderson	Simulatiestudie	Nee	14	Onvoldoende

De **acceptatie door patiënt** werd geëvalueerd op basis van privacyzorgen in twee kwalitatieve studie met voldoende methodologische kwaliteit (zie tabel). De bevindingen over privacy kwamen in één studie vanuit het perspectief van zorgverleners, en dus niet rechtstreeks van patiënten. In de andere studie werden wel interviews met patiënten afgenomen, maar betrof het een beperkt aantal (n=6).

Praktijkonderzoek naar de houding van patiënten, rechtstreeks met patiënten, ten aanzien van AI-spraak applicaties kan meer inzicht geven in de informatiebehoeften rondom SGR in de HAZ vanuit patiëntperspectief.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Shah_b	Kwalitatieve studie	Nee	22	Voldoende
Ha	Mixed-methodsstudie	Nee	6	Voldoende

De **adoptie door zorgprofessional** werd geëvalueerd op basis van gebruik, bruikbaarheid en toekomstig gebruik in zes studies; drie daarvan hadden een voldoende methodologische kwaliteit (zie tabel). Van de studies met voldoende methodologische kwaliteit werd in geen van de studies de houding van zorgverleners ten opzichte van AI-spraak applicaties vergeleken met reguliere verslaglegging. De steekproefgrootte in de studies met voldoende methodologische kwaliteit was groot, met tussen de 12 en 155 deelnemers.

Er is maar één studie met een redelijke steekproefgrootte en voldoende kwaliteit. Aanvullend onderzoek kan meer inzicht bieden in de adoptie van SGR in de HAZ.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Bundy	Kwalitatieve studie	Nee	12	Voldoende
Shah_a	Kwantitatieve studie	Ja	48	Onvoldoende
Shah_b	Kwalitatieve studie	Nee	22	Voldoende
Ha	Mixed-methodsstudie	Nee	155 zorgaanbieders (enquête), 11 zorgaanbieders (interviews)	Voldoende
Ma	Observationele studie	Nee	45	Onvoldoende
Van Buchem	Simulatiestudie	Nee	27	Onvoldoende

Werkplezier onder zorgverleners werd geëvalueerd in zes studies; vier daarvan hadden een voldoende methodologische kwaliteit (zie tabel). In de studies met voldoende methodologische kwaliteit werd het gebruik van AI-spraakapplicaties in twee studies vergeleken met reguliere verslaglegging. Het studiedesign van één van deze studies was van matige kwaliteit (niet-gerandomiseerde studie). De steekproefgrootte in die studie was redelijk, met 140 deelnemers.

Er lijkt voldoende onderzoek te zijn gedaan naar het werkplezier, aangezien de studie van voldoende kwaliteit met vergelijking en een redelijke steekproefgrootte.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Bundy	Kwalitatieve studie	Ja (kwalitatief)	12	Voldoende
Liu	Niet gerandomiseerde studie	Ja	140	Voldoende
Shah_a	Kwantitatieve studie	Ja	48	Onvoldoende
Shah_b	Kwalitatieve studie	Nee	22	Voldoende
Ha	Mixed-methodsstudie	Nee	155 zorgaanbieders (enquête), 11 zorgaanbieders (interviews)	Voldoende
Misurac	Observationele studie	Ja	35	Onvoldoende

De impact van AI-spraakapplicaties op **administratieve werkdruk** werd geëvalueerd op basis van de tijd voor verslaglegging, de tijd voor verslaglegging na werktijd (overwerk) en de bijdrage van zorgverleners aan de bewerking van AI-gegenereerde notities in negen studies; vier daarvan hadden een voldoende methodologische kwaliteit (zie tabel). In de studies met voldoende methodologische kwaliteit werd in drie studies het gebruik vergeleken met reguliere verslaglegging. De steekproefgrootte in twee van deze studies was redelijk, met 155 deelnemers en 140 deelnemers. Daarnaast was het studiedesign in één van de studies van matig niveau.

Er lijkt voldoende onderzoek te zijn gedaan naar de impact van AI-spraak applicaties op administratieve werkdruk, aangezien de studies van voldoende kwaliteit met vergelijking en een redelijke steekproefgrootte.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Bundy	Kwalitatieve studie	Ja (kwalitatief)	12	Voldoende
Biro _b	Simulatiestudie	Nee	11	Voldoende
Shah_a	Kwantitatieve studie	Ja	48	Onvoldoende
Shah_b	Kwalitatieve studie	Nee	22	Voldoende
Liu	Niet gerando-miseerde studie	Ja	140	Voldoende
Ma	Observationele studie	Ja	45	Onvoldoende
Ha	Mixed-methodsstudie	Ja	155 zorgaanbieders (enquête), 11 zorgaanbieders (interviews)	Voldoende
Van Buchem	Simulatiestudie	Ja	27	Onvoldoende
Hose	Simulatiestudie	Nee	11	Onvoldoende
Wang	Simulatiestudie	Ja	64	Onvoldoende

Het **perspectief van de zorgprofessional op de aandacht voor het gesprek** werd geëvalueerd in drie kwalitatieve studies; alle drie hadden een voldoende methodologische kwaliteit (zie tabel). In alle studies werd het perspectief van zorgverleners op de patiënt-zorgverlenerrelatie vergeleken met reguliere verslaglegging. De steekproefgrootte in deze studies was redelijk, met in totaal 189 deelnemers.

Er lijkt voldoende onderzoek te zijn gedaan naar het perspectief van zorgverleners op de patiënt-zorgverlenerrelatie, aangezien de studies van voldoende kwaliteit waren, een vergelijking maakten en een redelijke steekproefgrootte hadden.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Bundy	Kwalitatieve studie	Ja (kwalitatief)	12	Voldoende
Ha	Mixed-methodsstudie	Ja (kwalitatief)	155 zorgaanbieders (enquête), 11 zorgaanbieders (interviews)	Voldoende
Shah_b	Kwalitatieve studie	Ja (kwalitatief)	22	Voldoende

Het perspectief van de patiënt op de aandacht voor het gesprek werd geëvalueerd in twee studies met een voldoende methodologische kwaliteit (zie tabel). Eén studie had ook een design van gemiddelde bewijskracht. Het doel van deze studie was specifiek het evalueren van verschillen in patiënttevredenheid tussen consulten waarin een AI-spraak applicatie werd gebruikt voor verslaglegging en consulten met reguliere verslaglegging, met 288 patiënten in de open- en 304 in de geblindeerde fase. In de andere studie werden zes patiënten geïnterviewd over hun perspectief op de patiënt-zorgverlenerrelatie.

Er lijkt voldoende onderzoek te zijn gedaan naar het perspectief van patiënten op de patiënt-zorgverlenerrelatie.

(Eerste) auteur	Studiedesign	Vergelijking	Steekproefgrootte	Kwaliteitsbeoordeling*
Ha	Mixed-methodsstudie	Ja (kwalitatief)	6	Voldoende
Owens	Observationele studie	Ja	288 (open fase), 304 (geblindeerde fase)	Voldoende

Voor **begrijpelijkheid** zijn geen relevante uitkomsten gevonden. Begrijpelijkheid gaat over de mate waarin patiënten de automatisch gegenereerde SOEP-rapportage begrijpen en ervaren als duidelijk en concreet.

Praktijkonderzoek naar de begrijpelijkheid van AI-gegenereerde notities door patiënten kan meer inzicht geven in deze indicator.

Bijlage F – Toepassingen voor de HAZ

Toepassing	Digizo getoetst	HAZ	Claims	Herkomst data	Notitie
Autoscriber	Mee bezig	Ja	1 Bespaart tot 25% administratietijd per consultatie 2 De meerderheid (65%) meldt een toename in werk-tevredenheid 3 De gemiddelde gebruikersbeoor-deling voor de snelheid van samen-vatten bedraagt 94,7%, voor compleetheid 4,05/5 sterren en voor accuraatheid 3,85/5 sterren	1 Wetenschappelijk artikel LUMC 2 Samenvatting resultaten uit drie studies, waaronder artsen van DC clinics, WZA en ArboUnie Luo, G., Mao, S., Sinha, U., van Buchem, M. M., Kant, I. M. J., King, L., Kazmaier, J., Steyerberg, E. W., & Bauer, M. P. (2024). Impact of a Digital Scribe System on Clinical Documentation Time and Quality: Usability Study. Jmir Ai, 3. https://doi.org/10.2196/60020	Ambient listening (spraakgestuurd rapporteren tijdens gesprek) Ook achteraf (Smart Dictation) Site verwijst naar dit artikel: https://pmc.ncbi.nlm.nih.gov/articles/PMC6874666/
HealthTalk	Mee bezig	POH GGZ	1 Per consult een tijdsbesparing van 30% 2 In de voorbereiding van het integral overzicht van de cliënt nog een extra besparing van 4 uur bovenop • AI-gebaseerde verslaglegging bespaart 30% van de dagelijkse tijd en elimineert alle (vervelende) administratieve last	1 Tweejarig onderzoek van het Daisy project 2 Tweejarig onderzoek van het Daisy project	Spraakgestuurd rapporteren (achteraf) Ambient listening (spraakgestuurd rapporteren tijdens gesprek)
Juvoly	Ja	Ja	1 Zorgverleners besparen tijd, energie en ze registreren uniformer zodat ze meer kunnen betekenen voor hun patient • Het merendeel van de zorgverleners ervaart enige tot veel tijdswinst bij het gebruik van Juvoly QuickConsult tijdens het werk	1 Onderzoek in Rotterdam	Ambient listening (spraakgestuurd rapporteren tijdens gesprek) Spraakgestuurd rapporteren (achteraf)

Medendo	Ja	Ontworpen voor de gehele gezond-heidszorg en in het bijzonder de GGZ	1 Besparingen in de indirecte tijd / verslaglegging tot meer dan 50% van de oorspronkelijke tijd • Vlucht patiënt-ervaring (omdat de zorgverlener minder op het scherm hoeft te kijken)	1 Door de zorgverleners gerapporteerde resultaten	Spraakgestuurd rapporteren (achteraf) Ambient listening (spraakgestuurd rapporteren tijdens gesprek)
Tell James	Ja	Niet de focus en geen vermelding dat ze al actief zijn in HAZ maar wel geschikt voor HAZ	• Tijdswinst • Door direct rapporteren is 'het hoofd direct leeg' voor de volgende afspraak • De kwaliteit en leesbaarheid van de rapportages verhoogt	• Enquêtes • 0- en 1-metingen bij aangesloten zorgorganisaties	Spraakgestuurd rapporteren (achteraf) Ambient listening (spraakgestuurd rapporteren tijdens gesprek)
ValueCare	Ja	Ja	De Digitale Assistent verlaagt administratieve lasten en maakt zorg doelmatiger: 1 30% lastenvermin-dering/tijdswinst 2 Goede dataprivacy en wetgeving 3 Goede stabiliteit en performance 4 Goede Kwaliteit en leesbaarheid 5 Goede Helpdesk, training en proces-ondersteuning (beoordeling behandelaren) • Meer werkplezier • Meer direct klantcontact	1 Gemeten bij zorgprofessionals 2 ISO27001/NEN7510, VWO-BOZ-AI, ISAE3000, AI-act 3 Via dashboard 4 Beoordeling behandelaren 5 Beoordeling behandelaren	Spraakgestuurd rapporteren (achteraf) Ambient listening (spraakgestuurd rapporteren tijdens gesprek)
Verticai	Ja	POH GGZ	Geen harde claims		Spraakgestuurd rapporteren (achteraf) Ambient listening (spraakgestuurd rapporteren tijdens gesprek)