



# Klinisch oordeel versus statistische predictiemodellen: wie wint?

### Referentie

Colunga-Lozano LE, Foroutan F, Rayner D, et al. Clinical judgment shows similar and sometimes superior discrimination compared to prognostic clinical prediction models: a systematic review. *J Clin Epidemiol* 2024;165:111200. DOI: 10.1016/j.jclinepi.2023.10.016

### Duiding

Louise Joly, médecin généraliste, ULiège  
Geen belangenvermenging met het onderwerp.

## Klinische vraag

Wat is de voorspellende waarde van klinische predictiemodellen versus het klinisch oordeel van zorgprofessionals?

## Achtergrond

Het omgaan met diagnostische, therapeutische en prognostische onzekerheid is een van de grootste uitdagingen voor zorgprofessionals (1-3). Prognostische hulpmiddelen die gebaseerd zijn op statistische predictiemodellen (zoals SCORE-risico) kunnen zorgverleners helpen om de waarschijnlijkheid van het optreden van een gebeurtenis in te schatten (4). We bespraken in Minerva reeds het nut van bepaalde predictieve scores, echter zonder vergelijking met het klinisch oordeel (5-8). Ook het belang van de negatief voorspellende waarde van het niet-pluisgevoel van huisartsen bij de diagnose van kanker en ernstige ziekten kwam in Minerva reeds aan bod (9,10). Studies die de performantie van deze instrumenten vergeleken met het klinisch oordeel tonen tegenstrijdige resultaten en de weinige beschikbare systematische reviews beperkten zich voornamelijk tot acute zorg (4). De auteurs van de hier geduide systematische review stelden als doel voorop deze vergelijking uit te breiden naar alle predictieve instrumenten en, indien mogelijk, te verklaren waarom de ene boven de andere predictieve aanpak superieur is (4).

## Samenvatting

### Methodologie

Systematische review met kwantitatieve analyse (zonder pooling van de resultaten).

### Geraadpleegde bronnen

- Medline via Pubmed, Embase en CINAHL.

### Geselecteerde studies

- observationele studies die klinische predictiemodellen vergeleken met het klinisch oordeel
  - op dezelfde patiëntenpopulatie
  - gedurende dezelfde periode
  - en die één of twee metingen rapporteerden voor performantie, discriminerend of kalibratievermogen
- zonder taalrestricties
- men selecteerde in totaal 41 studies, gepubliceerd tussen 1986 en 2022; hiervan waren er 8 abstracts van congressen, 27 monocentrische studies ; de follow-up bedroeg 7 dagen tot 5 jaar.

### Bestudeerde populatie

- zorgprofessionals
- gehospitaliseerde en ambulante patiënten.

### Uitkomstmeting

- primaire uitkomstmaten
  - kenmerken van het gebruikte klinische predictiemodel (KPM) en de performantie qua discriminatie en kalibratie (vermogen om nauwkeurig te voorspellen wat er in werkelijkheid gebeurt)
  - performantie van het klinisch oordeel qua discriminatie en kalibratie
  - vergelijking van de performantie tussen het klinisch oordeel en een KPM
  - vergelijking van de performantie van het klinisch oordeel alleen versus het klinisch oordeel in combinatie met een KPM
- secundaire uitkomstmaten
  - subgroepanalyse:
    - type populatie
    - type klinische uitkomstmaat
    - zorgsetting (ziekenhuis of ambulant)
    - type KPM
    - type ontwikkeling KPM
    - type zorgverlener
    - monocentrisch/multicentrisch
    - steekproefgrootte
    - incidentie van gebeurtenissen
    - statistische performantie
- resultaten van primaire uitkomstmaten worden weergegeven in de vorm van ROC-curves met berekening van de area under the curve (AUC)
- toegevoegde waarde van het gebruik van het klinisch oordeel versus KPM.

### Resultaten

- primaire uitkomstmaten
  - globaal discriminerend vermogen: 31 studies

Type test	Mediane AUC	Interkwartiel bereik
KPM	0,73	0,62 tot 0,81
Klinisch oordeel	0,71	0,62 tot 0,81

- vergelijking discriminerend vermogen: de auteurs voerden 156 vergelijkingen uit
- KPM versus klinisch oordeel: 124 bruikbare vergelijkingen (28 studies)
  - KPM > klinisch oordeel: 31 vergelijkingen (15 studies)
  - klinisch oordeel > KPM: 35 vergelijkingen (13 studies)
  - geen significant verschil: 58 vergelijkingen (16 studies)

Type test	Mediane AUC	Spreiding (bovenste en onderste grenzen)
KPM	0,63	0,41 tot 0,86
Klinisch oordeel	0,66	0,59 tot 0,77

- klinisch oordeel versus klinisch oordeel + KPM: 30 bruikbare vergelijkingen (15 studies)
  - KPM + klinisch oordeel > klinisch oordeel: 9 studies
  - klinisch oordeel > KPM + klinisch oordeel: 1 studie

Type test	Mediane AUC	Spreiding (bovenste en onderste grenzen)
Klinisch oordeel	0,68	0,52 tot 0,89
KPM + klinisch oordeel	0,76	0,52 tot 0,88

- kalibratievermogen: weinig studies beschikbaar en bovendien heterogene gegevens
- goed kalibratievermogen voor KPM in 3 studies en voor klinisch oordeel in 1 studie
- in 4 studies zag men een beter kalibratievermogen voor KPM dan voor klinisch oordeel
- secundaire uitkomstmaten
  - subgroepanalyses:
    - klinisch oordeel i.v.m. prognose:
      - superieur bij verpleegkundigen versus huisartsen
      - superieur op spoedgevallendiensten versus diensten voor geestelijke gezondheidszorg
    - KPM i.v.m. prognose:
      - superieur in spoedgevallendiensten
      - discriminerend vermogen kwaliteitsvol en consistent voor alle populaties
      - beter discriminerend vermogen als KPM gebaseerd is op machine-learning dan op Bayesiaanse modellen
    - KPM versus klinisch oordeel:
      - KPM > klinisch oordeel bij chirurgische patiënten
      - klinisch oordeel > KPM bij patiënten op intensieve zorgen
      - klinisch oordeel = KPM in de acute zorg
    - de incidentie van gebeurtenissen is in de meeste studies laag (<100).

### **Besluit van de auteurs**

De auteurs concluderen dat het discriminerend vermogen van het klinisch oordeel minstens even valide is als dat van een klinisch predictiemodel (KPM), maar het kalibratievermogen van het klinisch oordeel is zwakker dan dat van een KPM.

### **Financiering van de studie**

Financiering door een particuliere stichting.

### **Belangenconflicten van de auteurs**

Geen belangenconflicten gemeld.

## **Bespreking**

### **Beoordeling van de methodologie**

Over het algemeen is deze studie van goede methodologische kwaliteit. Het protocol werd vooraf geregistreerd, de inclusie- en exclusiecriteria zijn duidelijk gerapporteerd. De literatuurzoektocht gebeurde in ten minste 3 databanken en een bibliothecaris was erbij betrokken. Twee auteurs voerden het selectieproces uit en kwamen hierin voor 85% overeen. De extractietabel is duidelijk en komt overeen met de onderzoeksvraag. De risico's van bias werden correct beoordeeld met behulp van de Prediction model Risk Of Bias ASsessment Tool (PROBAST). Voor 2 studies is er globaal een laag risico en voor 39 studies een hoog risico van bias. Risico van bias had vooral betrekking op de analyse van de resultaten waarbij er te weinig informatie over de statistische analyse gegeven werd en/of minder dan 100 gebeurtenissen (incidentie) voorkwamen. We kunnen echter niet voorbij aan een aantal zwakke punten van deze studie. Enerzijds bevat het protocol enkele onzorgvuldigheden die vragen oproepen over de manier waarop de analyses zijn uitgevoerd. Zo verschilt het aantal studies dat informatie rapporteert over het kalibratievermogen in verschillende paragrafen van 12 naar 13 en vervolgens naar 14. Daarnaast geven de auteurs aan dat hun analyse gebeurde met studies van 1986 tot 2022, terwijl enkele bladzijden verder wordt gesproken over studies van 1996 tot 2022. Dat zijn misschien wel details, maar op die manier rijzen twijfels over de nauwkeurigheid van de resultaten. Het valt op dat deze talrijke onvolkomenheden en benaderingen in de rapportage van de resultaten contrasteert met de nauwkeurigheid van het gepresenteerde protocol. Daarom interpreteren we de resultaten met grote voorzichtigheid en stellen we de herkomst van de gepresenteerde gegevens in vraag.

## Beoordeling van de resultaten van de studie

De onderzoeksvraag is nuttig voor de praktijk. Nochtans brengt de studie onder de noemer ‘klinisch predictiemodel’ (KPM) instrumenten samen die van toepassing zijn op een grote verscheidenheid aan klinische situaties, settings (CHA<sub>2</sub>DS<sub>2</sub>, Glasgow-score, enzovoort), zorgverleners met verschillende competenties, verantwoordelijkheden en ervaring. De auteurs erkennen de aanzienlijke heterogeniteit van de geïncludeerde studies. Dat maakt het moeilijk om de resultaten te veralgemenen. In vergelijking met andere studies bevestigt deze review echter wel de plaats van het klinisch oordeel als geldig prognostisch hulpmiddel (10-14).

## Wat zeggen de richtlijnen voor de klinische praktijk?

We zijn niet op de hoogte van praktijkrichtlijnen die een direct verband leggen tussen een klinisch predictiemodel (KPM) en het klinisch oordeel.

## Besluit van Minerva

Deze studie van goede methodologische kwaliteit met heterogene gegevens én met een globaal hoog risico van bias, suggereert dat het prognostisch vermogen van het klinisch oordeel van zorgverleners over het algemeen even accuraat is als dat van klinische predictiemodellen.

## Referenties

1. Motte B, Aiguier G, Van Pee D, Cobbaut JP. Mieux comprendre l’incertitude en médecine pour former les médecins. *Pédagogie médicale* 2020;21:39-51. DOI: 10.1051/pmed/2020025
2. Bloy G. L’incertitude en médecine générale : sources, formes et accommodements possibles. *Sciences sociales et santé* 2008;26:67. DOI: 10.3917/sss.261.0067
3. Higgs J, Jensen GM, Loftus S, Christensen N. (eds). *Clinical reasoning in the health professions*. Fourth edition. Elsevier, 2018.
4. Colunga-Lozano LE, Foroutan F, Rayner D, et al. Clinical judgment shows similar and sometimes superior discrimination compared to prognostic clinical prediction models: a systematic review. *J Clin Epidemiol* 2024;165:111200. DOI: 10.1016/j.jclinepi.2023.10.016
5. Henrard G. Heeft dit kind appendicitis? *Minerva* 2013;12(5):60-1.
6. Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol* 2013;66:95-104. DOI: 10.1016/j.jclinepi.2012.09.004
7. Van Meerhaeghe A. De waarde van een predictief model voor vroegtijdige diagnostiek van longkanker? *Minerva* 2013;12(9):112-3.
8. Iyen-Omofoman B, Tata LJ, Baldwin DR, et al. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013;68:451-9. DOI: 10.1136/thoraxjnl-2012-202348
9. Joly L. Wat is de diagnostische waarde van het buikgevoel van huisartsen bij het opsporen van kanker en ernstige aandoeningen? *Minerva* 2023;22(9):215-9.
10. Oliva-Fanlo B, March S, Gadea-Ruiz C, et al; CORap group. Prospective observational study on the prevalence and diagnostic value of general practitioners’ gut feelings for cancer and serious diseases. *J Gen Intern Med* 2022;37:3823-31. DOI: 10.1007/s11606-021-07352-w
11. Donker GA, Wiersma E, van der Hoek L, Heins M. Determinants of general practitioner’s cancer-related gut feelings - a prospective cohort study. *BMJ Open* 2016;6:e012511. DOI: 10.1136/bmjopen-2016-012511
12. Hartley MN, Sagar PM. The surgeon’s “gut feeling” as a predictor of post-operative outcome. *Ann R Coll Surg Engl* 1994;76(6 Suppl):277-8.
13. Oliva-Fanlo B, March S, Medina D, et al. Prevalence and diagnostic value of GPs’ gut feelings for cancer and serious diseases: protocol for a prospective observational study of diagnostic validity. *BMJ Open* 2019;9:e032404. DOI: 10.1136/bmjopen-2019-032404
14. Van den Bruel A, Thompson M, Buntinx F, Mant D. Clinicians’ gut feeling about serious infections in children: observational study. *BMJ* 2012;345:e6144-e6144. DOI: 10.1136/bmj.e6144