



## UvA-DARE (Digital Academic Repository)

### Decision support in hospital care for older patients

*Medication, falls and delirium*

Damoiseaux-Volman, B.A.

#### Publication date

2022

[Link to publication](#)

#### Citation for published version (APA):

Damoiseaux-Volman, B. A. (2022). *Decision support in hospital care for older patients: Medication, falls and delirium*. [Thesis, fully internal, Universiteit van Amsterdam].

#### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# 7

## Clinical validation of clinical decision support systems for medication review: a scoping review



Birgit A. Damoiseaux-Volman, Stephanie Medlock, Delanie M. van der Meulen, Jesse de Boer, Johannes A. Romijn, Nathalie van der Velde, Ameen Abu-Hanna

*Published in British Journal of Clinical Pharmacology, November 2021*

## **Abstract**

The aim of this scoping review is to summarize approaches and outcomes of clinical validation studies of clinical decision support systems (CDSSs) to support (part of) a medication review. A literature search was conducted in Embase and Medline. In total, 30 articles validating a CDSS were ultimately included. Most of the studies focused on detection of adverse drug events, potentially inappropriate medications and drug-related problems. We categorized the included articles in three groups: studies subjectively reviewing the clinical relevance of CDSS's output (21/30 studies) resulting in a positive predictive value (PPV) for clinical relevance of 4-80%; studies determining the relationship between alerts and actual events (10/30 studies) resulting in a PPV for actual events of 5-80%; and studies comparing output of CDSSs to chart/medication reviews in the whole study population (10/30 studies) resulting in a sensitivity of 28-85% and specificity of 42-75%. We found heterogeneity in the methods used and in the outcome measures. The validation studies did not report the use of a published CDSS validation strategy. To improve the effectiveness and uptake of CDSSs supporting a medication review, future research would benefit from a more systematic and comprehensive validation strategy.

## Introduction

Drug-related problems (DRP), such as adverse drug reactions (ADRs) and adverse drug events (ADEs), are common. For example, approximately 10% of the patients experience an ADR during hospital stay and approximately 20% of patients experience an ADE in outpatients settings.<sup>192,193</sup> ADRs are associated with prolonged hospital stay.<sup>194</sup> About 70% of the ADEs occurring in older hospitalized patients and between 16-42% of the ADEs occurring in outpatients are preventable.<sup>193,195</sup> Individualized medication reviews can prevent ADRs/ADEs. During medication reviews, prescribed medications are evaluated by reviewing their indication, interactions and appropriateness to optimize medication use and improve health outcomes.<sup>32</sup>

Clinical decision support systems (CDSSs) can support medication reviews and detect potentially inappropriate prescribing.<sup>77</sup> CDSSs can, for example, identify patients in need of a medication review, identify ADRs/ADEs or offer advice on how medications should be monitored or changed. Effectiveness of previous CDSS interventions supporting (part of) a medication review for hospitalized patients varied. CDSSs were mostly effective on process-related outcomes, but not yet on patient-related outcomes.<sup>33,196</sup>

An intervention with a CDSS is complex and must address the needs and wishes of future users. The Two-Stream Model and the GUIDES checklist describe factors that contribute to the success of CDSSs.<sup>36,37</sup> These factors include data quality, underlying clinical knowledge, and the presentation and quality of the output/advice. To ensure that the output of a CDSS is accurate, of high quality, and useful for clinical practice, the system should be tested and validated.<sup>197</sup> According to the World Health Organization, validation includes proving and documenting that a system leads to the expected results.<sup>198</sup> A medication-related CDSS validation strategy consists of four steps: 1) technical validation to check whether the CDSS functions as expected, 2) retrospective validation to review whether the output is clinically relevant, actionable and useful, 3) prospective validation before implementation in a real-life EHR to check whether the CDSS fits in the workflow (e.g., timing and frequency) and finally 4) post-implementation validation for continuous improvement.<sup>41,199</sup> The validation process can prevent irrelevant output for the clinical setting, alert fatigue or low user acceptance. Validation of a CDSS is, therefore, key for successful implementation.<sup>41,199</sup> To date, a systematic review was published in 2012 on the detection of ADEs, which found that the accuracy of ADE rules was reported in only 50% of the ADE detection system studies.<sup>200</sup>



The rule accuracy was often poor and varied due to multiple factors including differences in validation methods.<sup>200</sup> However, this 2012 systematic review focused solely on ADE detection systems and did not focus on other ways of clinical decision support for a medication review. Currently, an overview of published validation methods and outcomes, for the different CDSSs to support (part of) a medication review, is missing. A scoping review can be conducted to systematically identify literature, clarify concepts or identify knowledge gaps.<sup>201</sup>

Therefore, we conducted a scoping review with the aim of creating an overview of validated CDSSs supporting (part of) a medication review. Our secondary aim was to summarize and assess the methods and outcomes of the validation studies.

## **Methods**

A literature search (see Appendix 1 in the Supporting Information) combining the categories corresponding to the concepts of “CDSS”, “medication review”, and “validation” was conducted in Embase and Medline up to 9 August 2021. Search results were deduplicated in Endnote X9.<sup>202</sup> The references of the included articles were checked to find additional relevant articles. The study protocol was published on our departmental website (Technical Reports at the Department of Medical Informatics), and the completed PRISMA checklist for scoping reviews can be found in Appendix 2 in the Supporting Information.<sup>203</sup>

### Eligibility

Articles were screened based on title and abstract by one reviewer (D.M. or B.D.) using pre-specified eligibility criteria. Title and abstract screening were performed by using the Rayyan research app.<sup>204</sup> Studies were included if they described the clinical validation of a CDSS (with or without a prediction model) for medication review, for adults ( $\geq 18$  years), and were published in English in a peer-reviewed journal. After title and abstract screening, the full text of the remaining articles were also screened using the eligibility criteria. The full-text screening was performed by two researchers (D.M. and B.D.), independently. The results were compared, discussed and, if necessary, a third researcher (S.M.) was consulted. CDSS was defined as “any computer program designed to help healthcare professionals to make clinical decisions”.<sup>205</sup> The Pharmaceutical Care Network Europe defines medication review as “a structured evaluation of a patient’s medicines with the aim of optimizing medicines use and improving health outcomes. This entails detecting drug-related problems and recommending interventions”.<sup>32</sup>

We included CDSSs supporting any part of the medication review process. CDSSs during prescribing (order-entry pop-ups) were excluded. For clinical validation, we included all studies assessing the quality of the output of the CDSS.

### Data extraction and risk of bias assessment

Data extraction and assessment of the risk of bias were conducted by two researchers (D.M. and B.D.), independently. The results were compared, discussed and, if necessary, a third researcher (S.M.) was consulted. A template for data extraction was developed by J.d.B. and B.D. We extracted study and CDSS characteristics, validation methods, and all outcomes related to clinical validation. We used these validation methods and outcomes in a narrative synthesis. The outcomes of individual studies were rated on relevance and applicability on a scale of 1-3 by two researchers (D.M. and B.D.), independently. Furthermore, we assessed the compliance of the studies to the framework of Scheepers-Hoeks et al.<sup>41</sup> The references of the included studies were screened for additional articles. Risk of bias was assessed using an adapted version of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2).<sup>206</sup> Originally, this tool was intended for risk of bias assessment in studies determining diagnostic accuracy against a reference standard. We adjusted the signaling questions in the QUADAS-2 for the applicability to CDSS validation studies. We did not exclude studies with a high risk of bias in the narrative analysis.

## **Results**

### Study selection

Figure 7.1 shows the PRISMA flow diagram for article selection. The search in Medline and Embase identified 2434 articles, and search in additional sources identified nine further articles. After screening and eligibility assessment, 30 studies were included in this scoping review. In total, three of the included 30 articles were identified through additional sources (reference checking).



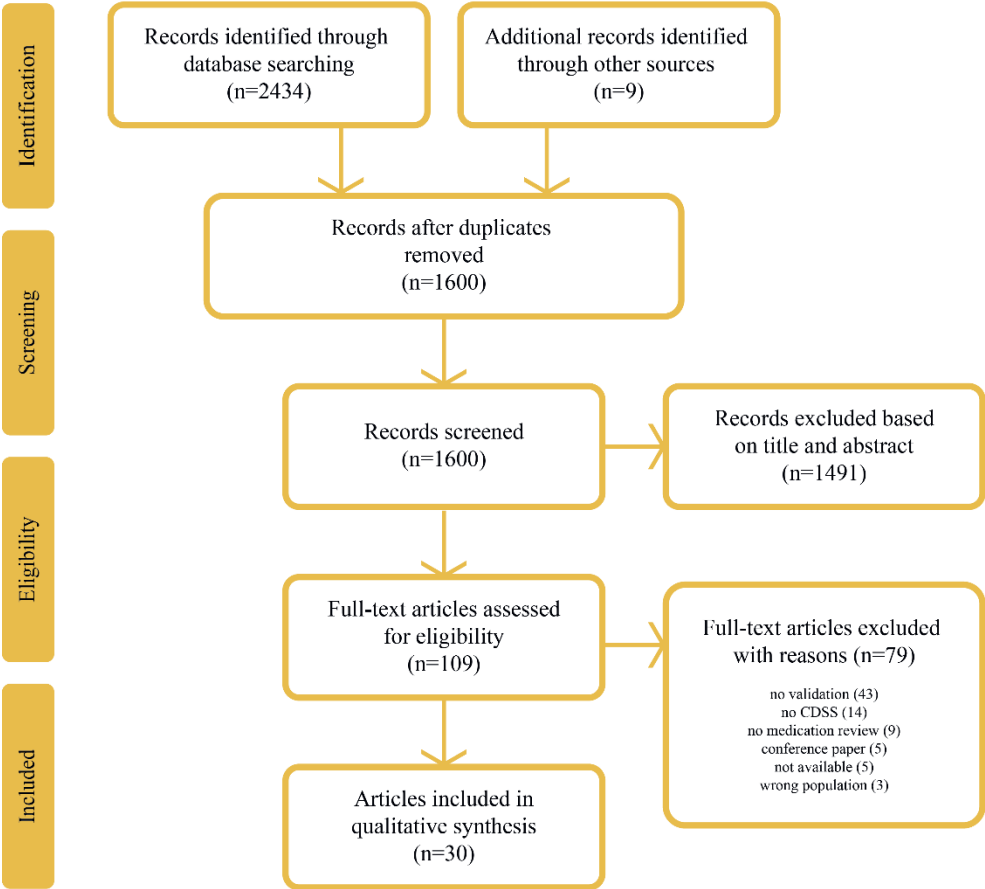


Figure 7.1. Prisma Flow Diagram

Characteristics

The included studies had been performed in 12 countries (Table 7.1): Belgium, Canada, Germany, Ireland, Israel, Korea, the Netherlands, Portugal, Spain, Sweden, Switzerland and the United States. Studies were published in the years 1998-2020. Of the 30 studies, 26 were conducted in or used data of an inpatient hospital setting, three of primary care, three of an outpatient (clinic), and two of a nursing home. The included studies used different clinical knowledge as input for their CDSSs. The alerts were based on literature, guidelines, experts' experience, prevalence and product guidelines.

In two studies, the CDSS was based on machine learning algorithms including statistical analysis and deep learning with neural networks and in one study the CDSS used an algorithm with risk scores based on literature.<sup>207–209</sup> Five of the 30 studies described the use of a tool for potentially inappropriate prescribing such as the Beers criteria, or STOPP/START criteria.<sup>210–214</sup> Of the included studies, 11 gave support to detect ADEs, six for potentially inappropriate medications (PIMs), five for DRPs, four for ADRs, three for medication surveillance, two for medication errors and two for drug-related hazardous conditions (DRHCs). The output of the CDSSs were presented as a list, (paper or electronic) report, database with alerts, an extra line in pharmacy module, alerts printed in the pharmacy or as a risk score.





Table 7.1. Study characteristics and CDSS description of the included studies

Authors	Study setting	Patients	CDSS description	CDSS input	CDSS output
Arvisais et al (2015) <sup>211</sup> Canada	Hospital, all (except 4) wards	n=182 (200 patient-days) Mean age=84 Female=36%	Computerized alert system (CAS), used by pharmacists	n=5 geriatric explicit criteria by expert panel, Beers criteria, prevalence	Alerts (on HTML page), daily generated by database
Azaz-Livshits et al (1998) <sup>215</sup> Israel	Hospital, medical ward	n=150 patients Mean age=NR Female=50%	ADR detection tool	n=14 ADR signals from laboratory data	List of alerts
Buckley et al (2018) <sup>216</sup> USA	Hospital, ICU and general ward	n=634 Mean age=60 n-ICU, 59 ICU Female=7% non-ICU, 42% ICU	DRHC triggers	n=20 unique trigger alerts	Paper and electronic reports
Cossette et al (2019) <sup>214</sup> Canada	Primary care	n=369 patients Mean age=77 Female=71%	Computerized alert system (CAS)	n=11 types of geriatric explicit criteria, 2015 Beers, STOPP v2	List of patients with PIMs and alert reasons
Dalton et al (2020) <sup>212</sup> Ireland	Hospital	n=204 patients Mean age=77 Female=49%	SENATOR software, PIM CDSS for physicians	n=111 recommendations based on STOPP/START v2	Report per patient on paper and per email
DiPoto et al (2015) <sup>217</sup> USA	3 hospitals, ICU, general ward	n=623 Mean age=60 n-ICU, 58 ICU Female=49% n-ICU, 39% ICU	Automated surveillance trigger alerts for DRHCs	n=93 logic based trigger rules	Alerts in centralized database
Dormann et al (2000) <sup>218</sup> Germany	Hospital, medical ward	n=379 patients age=51 Female=35%	Computer-based ADR monitoring system	n=20 automated laboratory signals	Daily list of alerts with patient name, date of event
Eppenga et al (2012) <sup>219</sup> Netherlands	Hospital	n=619 patients Mean age=53 Female=54%	2 medication surveillance CDSSs: 1) Centrasys (iSOFT) 2) Pharmaps advanced	n=NR, CDSS 1 = G-standard (Dutch drug database), CDSS 2 = G-standard + expert based rules	CDSS 1: alerts at one point of time CDSS 2: generation of alerts once a day
Ferrandez et al (2017) <sup>220</sup> Spain	Hospital	n=17,878 Mean age=55 Female=49%	DRP alert in pharmacy warning system, used by pharmacists	n=414 medications (7879 alerts), based on literature, selected by two pharmacists	Extra line in pharmacy module. Alerts include patient, dose and recommendations.

Table 7.1. (continued)

Authors	Study setting	Patients	CDSS description	CDSS input	CDSS output
Fritz et al (2012) <sup>221</sup> Switzerland	Hospital, 2 general internal wards	n=100 Median age =59 Female=42%	3 CDSS medication surveillance: 1) Pharmavista 2) DrugReax 3) TheraOpt	NR	Alerts
Garcia-Caballero et al (2018) <sup>213</sup> Spain	Nursing home	n=115 patients Mean age=79 Female=62%	PIP Screening (Polimedicamentation)	n=NR, STOPP criteria	Automated alerts
Hammar et al (2015) <sup>222</sup> Sweden	2 geriatric clinics, 3 primary care units	n=254 patients Mean age=84 Female=64%	Electronic expert support system (EES) for DRPs, used by physicians	n=NR, DDIs based on Swedish databank	Paper based reports with potential DRPs
Hedna et al (2019) <sup>209</sup> Sweden	110 clinics 51 primary care units, 29 departments, 3 hospitals	n=745 patients Mean age=75 Female=57%	PHARAO, risk scores for ADE	n=9 (p)ADE, algorithm, database of potential interactions	Low, intermediate and high risk with advice
Hwang et al (2008) <sup>223</sup> Korea	Hospital, 2 ICUs, 5 general wards	n=598 patients Age=NR Female=NR	ADE monitor, integrated in HIS	n=46, based on previously studied CDSS in literature, modified by panel	List of alerts with report including medication, ADE, location
Ibanez-Garcia et al (2019) <sup>224</sup> Spain	Hospital	n=25,449 admissions Age=NR Female=NR	ADE CDSS (HIGEA)	n=211 clinical rules based on bibliographic search, clinical practice, team	Real-time list of alerts
Jha et al (1998) <sup>225</sup> USA	Hospital, 9 units	n=21,964 patient-days Age=NR Female=NR	ADE monitor	n=52 ADE detection rules based on literature, study team	List of alerts with report including name, bed, event, condition
Jha et al (2008) <sup>226</sup> USA	Hospital, 5 units	n=2407 patients Av. age=74 Female=53%	Dynamic Pharmacovigilance	n=NR, rule base, using data of laboratory, medication, patient demographic	List of alerts



Table 7.1. (continued)

Authors	Study setting	Patients	CDSS description	CDSS input	CDSS output
Levy et al (1999) <sup>227</sup> Israel	Hospital, medical ward	n=192 Age=~75% >60 Female=47%	ADR detection tool	n=NR, ADR signals from laboratory data (same tool as Azaz-Livshits et al)	List of alerts
Miguel et al (2013) <sup>228</sup> Portugal	Hospital	n=118 patients Mean age=60 Female=41%	ADR detection tool, stand-alone (patient data needs to be filled in by hand)	n=10 ADRs selection drugs and ADRs based on hospital formulary, INFARMED, book	Alerts with suggested ADRs and frequent ADR for prescribed drugs
Peterson et al (2014) <sup>210</sup> USA	Hospital, general medicine, orthopaedics, urology	n=179 patients Mean age=72 Female=NR	PIM review dashboard	n=NR, triggers based on Beers and STOPP criteria, anticholinergic risk scale, formulary	List of patients with PIM(s) sorted by highest risk
Quintens et al (2019) <sup>229</sup> Belgium	Hospital	n=NR Age=47-74 Female=NR	Check of Medication Appropriateness (CMA)	n=78 clinical rules for PIMs, DRPs and ADEs based on literature and guidelines, selected by team	Once a day generation of list of alerts on a worklist
Raschke et al (1998) <sup>230</sup> USA	Hospital, Nonobstetrical patients	n=9306 admissions Age=NR Female=NR	ADE system	n=37 ADE rules defined by authors	Alerts printed in pharmacy
Rommers et al (2011) <sup>231</sup> Netherlands	Hospital, general internal ward	n=NR Age=NR Female=NR	ADE alerting system (ADEAS), pharmacists	n=121 clinical expert based rules	Every morning, list of alerts
Rommers et al (2013) <sup>232</sup> Netherlands	Hospital, 6 wards	n=931 patients Age=NR Female=NR	ADE alerting system (ADEAS), pharmacists	n=121 clinical expert based rules	List of alerts for patients with possible ADE
Roten et al (2010) <sup>233</sup> Switzerland	Hospital, internal medicine and geriatric wards	n=501 patients Age=NR Female=NR	DRP screening tool	n=6 DRP queries based on literature, pharmacists, another hospital	List of patients with possible DRPs

Table 7.1. (continued)

Authors	Study setting	Patients	CDSS description	CDSS input	CDSS output
Schiff et al (2017) <sup>208</sup> USA	Outpatients	n=NR Age=NR Female=NR	Medication errors, Outlier detection screening (MedAware)	n=1706 alerts, Machine learning algorithms identifying outliers	Alerts with short explanation
Silverman et al (2004) <sup>234</sup> USA	Hospital	n=NR Age=NR Female=NR	ADE detection system	n=NR, ADE detection rules (modified version of Jha et al)	List of alerts
Segal et al (2019) <sup>207</sup> Israel	Hospital, internal medicine department	n=3160 patients Age=NR Female=NR	Medication errors / ADE CDSS (MedAware)	n=NR, machine learning algorithms identifying outliers	Alerts after change in clinical state patient
de Wit et al (2015) <sup>235</sup> Netherlands	Hospital Nursing home	n=900 patients Age=NR Female=NR	Medication surveillance CDSS, stand-alone, pharmacists	n=39 clinical rules based on product info, known ADEs, prescribing mistakes	Alerts
de Wit et al (2016) <sup>236</sup> Netherlands	Hospital, geriatric ward	n=33 Mean age=83 Female=45%	Medication review CDSS, stand-alone, used by pharmacists	n=469 clinical rules based on literature, guidelines, protocols, multidisciplinary team	DRP alerts with advice to prevent ADE

ADE = Adverse Drug Event, ADR = Adverse Drug Reaction, CDSS = Clinical Decision Support System, DRHC = Drug-Related Hazardous Condition, DRP = Drug Related Problem, ICU = Intensive Care Unit, NR = Not Reported, PIM = Potentially Inappropriate Medication, PIP = Potentially Inappropriate Prescribing

### Validation methods

All included studies used real-life patient data for the validation of the CDSSs. The studies included between 33 and 17,878 patients. Four studies did not report information on the number of patients or admissions.<sup>208,229,231,234</sup> Twenty-one (70%) studies reported how many (type of) rules were included in the knowledge base of the CDSS, ranging from 5 to 7879 alert (types).



As table 7.2 shows, all studies reported a frequency of generated alerts either as number of alerts or as number of patients/admissions with an alert. Of the 30 articles, eight studies used retrospective data for the validation (step 2 of the validation strategy of Scheepers-Hoeks et al), 17 studies used prospective data (step 3), and five studies used post-implementation data for continuous improvement (step 4). One study assessed the CDSS with a pre-defined outcome threshold and aimed for a sensitivity of  $\geq 80\%$ .<sup>233</sup> We divided the included studies into group A, B and/or C (Table 7.2). Group A assessed clinical validation by subjectively checking the clinical relevance of the alerts (21 studies). Group B checked the association of alerts/advice with actual occurrence of events (10 studies) and group C compared the CDSS's output with a chart/medication review in the whole study population (10 studies).

#### A - Clinical relevance of CDSS's output

In 21 of 30 studies, a reviewer checked the output of the CDSS and assessed the clinical relevance through their clinical experience.<sup>207,208,210-214,217,219-222,224,226,230-232,234-237</sup> The reviewers assessed the clinical relevance with or without checking the charts or contacting a physician/nurse. In 13 of the 21 studies, the reviewer was a pharmacist(s), in two studies a combination of a pharmacist and a physician, in one study a combination of a pharmacist and a radiology technician, in two studies a physician and in 3 studies the profession of the reviewer was not described.

#### B - CDSS's output and occurrence of actual events (in patients with alert)

In 10 of 30 studies, a reviewer checked the patient charts to find a relationship between the output of the CDSS and actual occurrence of an ADR/ADE/DRHC.<sup>209,215-218,223,225-228</sup> In eight of the 10 studies, all included patients with a CDSS alert were reviewed and in two studies, a sample of patients was reviewed.

C - CDSS's output and chart/medication review in the whole population (in patients with or without alert)

In 10 of 30 studies, the clinical validation was assessed by comparing the output of the CDSS with the output of a chart or medication review in the whole population (not only the patients with a CDSS alert).<sup>209,215,218,220,223,225,227,228,233,236</sup> In four of the 10 studies, a pharmacist conducted a medication review to find DRPs/ADEs. In four other studies, a team or clinical pharmacologist reviewed the charts for DRPs/ADEs/ADRs and two studies additionally used voluntary ADE/ADR reports. The reviewers were blinded to the output of the CDSS in four of the 10 studies.<sup>218,225,233,236</sup>

### Outcome measures and outcomes

A - Clinical relevance of CDSS's output

Table 7.2 shows that the 21 studies assessing the clinical relevance of CDSS's output used different names to indicate positive predictive value (PPV) for clinical relevance: (clinical) relevance, clinical usefulness, clinical validity, efficiency, PPV and rule effectiveness. We used the term PPV(clinical relevance) for all studies.

Twenty studies reported the PPV(clinical relevance) of alerts, using a total of 89,974 alerts ranging from 53 to 39,481 alerts (median=833 alerts, IQR=318-2639 alerts). A range of 4-80% of alerts were found to be clinically relevant (median=13%, IQR=8%-44%). One study reported the PPV(clinical relevance) in patient-days, finding that 75% of patient days had  $\geq 1$  clinically relevant alert.<sup>211</sup>

B - CDSS's output and occurrence of actual events (in patients with alert)

All 10 studies reported the proportion of alerts related to a possible ADE/ADR/DRHC as an outcome.<sup>209,215-218,223,225-228</sup> Eight of the 10 studies referred to this outcome as the PPV (Table 7.2). We used the term PPV for all studies. The PPV for ADEs varied between 5 and 23% (median=19%, IQR=14%-22%).<sup>216,223,225,226</sup> The PPV for ADRs varied between 13 and 80% (median=28%, IQR=17%-48%).<sup>215,218,227,228</sup> The PPVs for DRHCs were 29% and 71%.<sup>216,217</sup>



C - CDSS's output and chart/medication review in the whole population (in patients with or without alert)

All 10 studies reported the sensitivity (Sn), or related measure, as the proportion of the DRPs/ADRs/ADEs found by CDSSs.<sup>209,215–218,223,225–228</sup> The Sn for DRPs were 28%, 79% and 85%.<sup>220,233,236</sup> The methods to detect ADRs varied and the reported outcomes were therefore not comparable: Sn=62% (40/65 admissions, ADR detection by multiple methods), Sn=66% (25/38 admissions, by expert team), Sn=74% (34/46 ADRs, multiple methods), Sn=83% (10/12 ADRs, by chart review).<sup>215,218,227,228</sup> The Sn for ADEs were 45% and 79%.<sup>223,225</sup> Of the 10 studies, 5 reported the specificity (Sp).<sup>209,215,218,227,233</sup> The Sp for DRPs was 60% and the Sp for ADRs was 42–75%.

Table 7.2. Validation methods and outcomes of the included studies

Authors	Group A/B/C and description methods	Outcomes, relevance (R), applicability (A)			Compliance framework <sup>41</sup> step 2,3,4 & outcomes			
		R	A	2	3	4	Outcomes	
Arvisais et al (2015) <sup>211</sup>	A Two junior pharmacists (independently) analysed alerts for clinical relevance	Clinical relevance=75% (149 with ≥1 clinically relevant alert/ 200 patient days)	+	++ +	–	X	–	PPV(clinical relevance)=yes NPV=no Extra=no
Azaz-Livshits et al (1998) <sup>215</sup>	B Signals evaluated by expert team	37% (78 signals related to ADR/212 signals) 29% (25 admissions ADR & signal/86 admissions with alert)	++	++	X	–	–	PPV(clinical relevance)=no NPV=no Extra=yes
	C Clinical pharmacist reviewed charts for ADRs, evaluated by expert team	Sn=66% (25 CDSS/38 admissions by expert team) Sp=49% (56/115 admissions)	++ +	+				
Buckley et al (2018) <sup>216</sup>	B Pharmacist reviewed alerts and patient charts to determine causality for DRHCs/ADEs using validated tools	PPV (DRHCs)=29% (249/870 alerts) PPV (ADEs)=5% (47/870)	++	++	–	–	X	PPV(clinical relevance)=no NPV=no Extra=yes

Table 7.2. (continued)

Authors	Group A/B/C and description methods		Outcomes, relevance (R), applicability (A)			Compliance framework <sup>41</sup> step 2,3,4 & outcomes			
			R	A	2	3	4	Outcomes	
Cossette et al (2019) <sup>214</sup>	A	Clinical relevance assessment by 2 pharmacists based on clinical experience	Clinical relevance =41% (34/83 alerts) 42% (27/65 patients with ≥1 alert)	+	++ +	–	X	–	PPV(clinical relevance)= yes NPV=no Extra=no
Dalton et al (2020) <sup>212</sup>	A	Independent analysis by pharmacist and physician (6-point scale)	Clinical relevance=74% (681 relevant/925 alerts)	+	++ +	X	–	–	PPV(clinical relevance)= yes NPV=no Extra=no
DiPoto et al (2015) <sup>217</sup>	A	Pharmacists assessed alerts (clinical relevant: pharmacists proposed change)	Clinical relevance=40% (ICU, 90/226 alerts), 45% (general ward, 235/525 alerts)	+	++ +	–	–	X	PPV(clinical relevance)= yes NPV=no Extra=yes
	B	Pharmacist assessed causality trigger and adverse events (subgroup analysis of 161 triggers of 19 rules)	PPV (DRHCs)=71% (115/161 triggers)	++	++				
Dormann et al (2000) <sup>218</sup>	B	Team of pharmacologist, clinician and pharmacists assessed alerts and patient chart for ADR (Naranjo score)	PPV (ADRs)=13% (63/501 alerts)	++	++	–	X	–	PPV(clinical relevance)= no NPV=no Extra=yes
	C	Compare ADRs detected by CDSS with ADRs from spontaneous reporting	relative Sn=74% (34 ADRs CDSS/46 (all) ADRs) relative Sp=75%	+	+				
Eppenga et al (2012) <sup>219</sup>	A	Two pharmacists independently assessed alerts of 2 CDSSs for clinical relevance (pharmacist would take action)	PPV (clinical relevance) CDSS 1=6% (150/2607 alert), PPV (clinical relevance) CDSS 2=17% (384/2256 alerts)	+	++ +	–	X	–	PPV(clinical relevance)= yes NPV=no Extra=no





Table 7.2. (continued)

Authors	Group A/B/C and description methods	Outcomes, relevance (R), applicability (A)	Compliance framework <sup>41</sup> step 2,3,4 & outcomes						
			R	A	2	3	4	Outcomes	
Ferrandez et al (2017) <sup>220</sup>	A Alerts reviewed for clinical relevance (action needed) by pharmacists	Clinical relevance=20% (2808/13,833 alerts)	+	++ +	-	-	X	PPV(clinical relevance)=yes NPV=no Extra=yes	
	C Compare DRPs detected by CDSS with pharmacist review	79% (2808 DRPs by CDSS/3552 (all) DRPs)	++ +	+					
Fritz et al (2012) <sup>221</sup>	A Alerts reviewed for clinical relevance (action needed) by pharmacists AND sensitivity to detect all 33 relevant alerts (identified by pharmacist while reviewing alerts)	PPV(clinical relevance)=6%, 8%, 8% (3/53, 29/364, 25/328 alerts) Sn=9%, 88%, 76% (3/33, 29/33, 25/33 relevant alerts)	++	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=yes	
Garcia-Caballero et al (2018) <sup>213</sup>	A A physician and psychiatrist reviewed alerts	Relevance=12% (140/1155 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no	
Hammar et al (2015) <sup>222</sup>	A A physician reviewed alerts for clinical relevance as part of medication review	Clinical relevance=68% (502/740 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no	
Hedna et al (2019) <sup>209</sup>	B For each alert was determined whether it was related to a symptom	PPV=0.20-0.25 (low risk: 150/776, intermediate risk: 93/460, high risk: 53/208 alerts)	++	++	X	-	-	PPV(clinical relevance)=no NPV=yes Extra=yes	
	C Pharmacists extracted symptoms associated with medications, checked by second reviewer	Sn=0.12-0.37 (high-low, patients symptom & alert/patients symptom from review) Sp=0.78-0.95 (low-high) NPV=0.89-0.90 (high-low)	++	+					

Table 7.2. (continued)

Authors	Group A/B/C and description methods	Outcomes, relevance (R), applicability (A)	Compliance framework <sup>41</sup> step 2,3,4 & outcomes						
			R	A	2	3	4	Outcomes	
Hwang et al (2008) <sup>223</sup>	B Pharmacist reviewed alerts, charts for ADE association	PPV(ADEs)=21% (148/718)	++	++	X	-	-	PPV(clinical relevance)=no NPV=no Extra=yes	
	C Pharmacist (checked by 5 pharmacists) reviewed charts patients without alert for ADEs	Sn=79% (148/187 ADEs)	+++	+					
Ibanez-Garcia et al (2019) <sup>224</sup>	A Pharmacist reviewed alerts, and advised physician	51% (554 with advice/1086 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no	
Jha et al (1998) <sup>225</sup>	B Reviewer analysed ADE association (checked by physician)	PPV(ADEs)=17% (450/2620 alerts)	++	++	-	X	-	PPV(clinical relevance)=no NPV=no Extra=yes	
	C Reviewers (blinded to CDSS) conducted ADE detection study (3 methods)	ADEs detected by CDSS =45% (275/675 (all) ADEs)	+++	+					
Jha et al (2008) <sup>226</sup>	A Reviewer analysed 52% alerts, contacted physician if necessary	Clinical relevance=11% (30 with contact/266 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=yes	
	B Chart review for (potential) ADEs in a sample of patients (checked by physician)	PPV(ADEs)=23% PPV(pADEs)=15%	++	++					
Levy et al (1999) <sup>227</sup>	B Analyses of signals	18% (Signals related to ADR(52)/ all signals(295))	++	++	-	X	-	PPV(clinical relevance)=no NPV=no Extra=yes	
	C Team reviewed charts for ADRs	Sn=62% (40 ADR admissions by tool/65 (all) ADR admissions) Sp=42% (79/135 admission without ADR)	+++	+					



Table 7.2. (continued)

Authors	Group A/B/C and description methods	Outcomes, relevance (R), applicability (A)	Compliance framework <sup>41</sup> step 2,3,4 & outcomes					
			R	A	2	3	4	Outcomes
Miguel et al (2013) <sup>228</sup>	B ADRs detected by CDSS reviewed for true ADRs	PPV=80% (65 true ADRs/81 all suggested ADRs)	++	++	X	-	-	PPV(clinical relevance)=no NPV=no Extra=yes
	C Chart review and assessment by CDSS in population	83% (10 ADRs CDSS/12 ADRs in chart review)	+	++				
Peterson et al (2014) <sup>210</sup>	A Pharmacist reviewed patients on dashboard and advised physician	12% (22 with intervention/179 patients) 6% (31 with interventions/485 alerts (PIMs))	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no
Quintens et al (2019) <sup>229</sup>	A Pharmacist checked alerts (clinical relevance = action towards physician)	Clinical relevance=8% (3205 with action/ 39,481 alerts)	+	++ +	X	-	-	PPV(clinical relevance)=yes NPV=no Extra=no
Raschke et al (1998) <sup>230</sup>	A Pharmacist /radiology technicians evaluated alerts and advised physician	Relevance=71% (794 with advice/1116 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no
Rommers et al (2011) <sup>231</sup>	A Hospital pharmacists reviewed alerts	Clinical relevance =19% (14 with intervention/ 72 true positive alert)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no
Rommers et al (2013) <sup>232</sup>	A Pharmacists reviewed alerts, contacted and advised physician/nurse	PPV(clinical relevance)=8% (204 with advise/2650 alerts)	+	++ +	-	X	-	PPV(clinical relevance)=yes NPV=no Extra=no
Roten et al (2010) <sup>233</sup>	C Pharmacists conducted medication review (blinded to CDSS) to identify DRPs	324 (65%) patients with alert, Sn=85% (235 by CDSS/276 (all) patients with DRP), Sp=60% (136/225 (all) patients without DRP)	++ +	+	-	X	-	PPV(clinical relevance)=no NPV=no Extra=yes
Schiff et al (2017) <sup>208</sup>	A Chart of patients with an alert were reviewed for accuracy and clinical validity	126 alerts: Accuracy=93% (based on data) Clinical validity (relevance )=75%	+	++	X	-	-	PPV(clinical relevance)=yes NPV=no Extra=no

Table 7.2. (continued)

Authors	Group A/B/C and description methods	Outcomes, relevance (R), applicability (A)	Compliance framework <sup>41</sup> step 2,3,4 & outcomes					
			R	A	2	3	4	Outcomes
Silverman et al (2004) <sup>234</sup>	A Pharmacists reviewed alerts, and advised physician (3x with different ADE rules)	Rule effectiveness (Clinical relevance)= 5%, 6%, 13% (169/3117, 452/7390, 792/6136 alerts)	+	++ +	-	-	X	PPV(clinical relevance)= yes NPV=no Extra=no
Segal et al (2019) <sup>207</sup>	A Biweekly interviews to manually review alerts	315 alerts: Accuracy=89% (no data issues), Clinical validity=85% (no justification for medication), Clinical usefulness (Clinical relevance) =80%	+	++	-	X	-	PPV(clinical relevance)= yes NPV=no Extra=no
De Wit et al (2015) <sup>235</sup>	A Pharmacists reviewed alerts (clinical relevance =advised physician)	Efficiency(Clinical relevance)=4% (147/4065 alerts)	+	++ +	-	-	X	PPV(clinical relevance)= yes NPV=no Extra=no
de Wit et al (2016) <sup>236</sup>	A Pharmacist and geriatrician independently checked DRPs by CDSS	Clinical relevance= 12% (70/574 alerts) Sn=72.9% (51 classified as relevant by CDSS / 70 relevant alerts) Sp=98.6% (497 classified as irrelevant by CDSS / 504 irrelevant alerts)	+	++	X	-	-	PPV(clinical relevance)= yes NPV=no Extra=yes
	C Geronto-pharmacology meeting discussed DRPs from a medication review (blinded to CDSS)	20% (44 DRPs CDSS/223 DRPs medication review) 28% (70 DRPs CDSS/249 (all) DRPs)	++	+				

ADE = Adverse Drug Event, ADR = Adverse Drug Reaction, CDSS = Clinical Decision Support System, DRHC = Drug-Related Hazardous Condition, DRP =Drug Related Problem, group A = studying clinical relevance of CDSS's output, group B = CDSS's output and actual occurrence of DRPs (patients with alert), group C = CDSS's output and chart/medication review in whole population, PIM = Potentially Inappropriate Medication, PPV = Positive Predictive Value, Sn = sensitivity, Sp = specificity



Risk of bias assessment

Most criteria of the assessment were categorized as low or unclear risk of bias for the included studies (Table 7.3). Twenty studies did not use a reference standard (chart/medication review for the whole population) and therefore the category was not applicable.<sup>207,208,210–214,216,217,219,221,222,224,226,230–232,234,235,237</sup> Two studies scored a high risk of bias for one category.<sup>223,233,236</sup>

Table 7.3. Risk of bias and applicability concerns of the included studies

Authors	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow & timing	Patient selection	Index test	Reference standard
Arvisais et al (2015)	-	-	N/A	N/A	-	-	N/A
Azaz-Livshits et al (1998)	-	?	?	-	-	-	-
Buckley et al (2018)	-	-	N/A	N/A	-	-	N/A
Cossette et al (2019)	-	-	N/A	N/A	-	-	N/A
Dalton et al (2020)	-	-	N/A	N/A	-	-	N/A
DiPoto et al (2015)	-	-	N/A	N/A	-	-	N/A
Dormann et al (2000)	-	-	-	-	-	-	-
Eppenga et al (2012)	-	-	N/A	N/A	-	-	N/A
Ferrandez et al (2017)	-	-	-	?	-	-	-
Fritz et al (2012)	-	-	N/A	N/A	-	-	N/A
Garcia-Caballero et al (2018)	?	-	N/A	N/A	-	-	N/A
Hammar et al (2015)	-	-	N/A	N/A	-	-	N/A
Hedna et al (2019)	-	-	-	?	-	-	-
Hwang et al (2008)	-	-	+	-	-	-	-
Ibanez-Garcia et al (2019)	-	-	N/A	N/A	-	-	N/A
Jha et al (1998)	-	-	-	-	-	-	-
Jha et al (2008)	-	-	N/A	N/A	-	-	N/A
Levy et al (1999)	-	?	?	-	-	-	-

Table 7.3. (continued)

Authors	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow & timing	Patient selection	Index test	Reference standard
Miguel et al (2013)	-	-	-	?	-	-	-
Peterson et al (2014)	-	-	N/A	N/A	-	-	N/A
Quintens et al (2019)	-	-	N/A	N/A	-	-	N/A
Raschke et al (1998)	-	-	N/A	N/A	-	-	N/A
Rommers et al (2011)	-	-	N/A	N/A	-	-	N/A
Rommers et al (2013)	-	-	N/A	N/A	-	-	N/A
Roten et al (2010)	?	-	-	-	-	-	-
Schiff et al (2017)	-	-	N/A	N/A	-	-	N/A
Silverman et al (2004)	?	-	N/A	N/A	-	-	N/A
Segal et al (2019)	-	-	N/A	N/A	-	-	N/A
De Wit et al (2015)	-	-	N/A	N/A	-	-	N/A
de Wit et al (2016)	+	-	-	-	-	-	-

- Low, + High, ? Unclear

## Discussion

We found 30 articles describing the clinical validation of a CDSS to support (part of) a medication review. Most of the studies focused on detection of ADEs, PIMs and DRPs. The methods used in these 30 articles can be categorized into three groups: A) studies subjectively assessing the clinical relevance of CDSS's output (21/30), B) studies describing the association between alerts and actual events (10/30), and C) studies comparing CDSS's output against chart/medication review of the whole population (10/30).

In our scoping review, we found heterogeneity in the methods used and in the outcome measures. In alignment with our results, the systematic review from 2012 also found heterogeneity in the methods to assess the accuracy of ADE computerized detection rules.<sup>200</sup> We, therefore, divided the studies into three groups.



The studies in group A and B assessed the quality of the alerts (useful/correct) for patients with an alert. Studies in group A assessed the quality of the alerts by clinical experience of reviewers. This is a relatively easily applicable method and effective as a first step in assessing the CDSS validity. Studies in group B assessed the quality of the alerts through determining their association with medication errors (DRHCs, ADEs, ADRs). This increased the validation efforts of the reviewers but also increased the quality and relevance of the assessment. The studies in group C assessed the quality of the alerts by studying patients with and without an alert (coverage). The medication errors were found in the whole population through a medication review (or chart review, or spontaneous reporting) and these were compared with the output of the CDSSs. This is the most comprehensive and informative method as it gives insights into which medication errors the CDSS captured and which it missed. The outcome measures of group A are included in the framework of Scheepers-Hoeks et al, but the reported methods and outcomes of group B and C are not (yet) fully included in the framework.<sup>41</sup>

The PPV(clinical relevance) of alerts reported in our review was 4-80% with a median of 13%. This median clinical relevance is low when compared to the desired threshold of  $\geq 89\%$  mentioned in the framework of Scheepers-Hoeks et al.<sup>41</sup> The PPV(clinical relevance) reported in the studies of our review were calculated using retrospective and prospective data, based on 53 to 39,481 generated alerts, 33 to 17,878 patients, 5 to 7879 number of rules, and different kinds of support (ADE detection, medication surveillance, PIMs, DRPs, DRHCs and medication errors). The two studies using machine learning as input (knowledge base) had a relatively high PPV(clinical relevance) of 75% and 80%. We did not find another variable or pattern explaining the difference between low vs. high proportion of relevant alerts. The studies offered commentary on the reasons for low/high clinical relevance and for the impact of the clinical relevance. The clinical relevance would improve with higher data quality, by including more patient characteristics in the CDSSs and by the selection of high impact or prevalent problems.<sup>211,219,224,225,229</sup> In the study of Dalton et al, a higher clinical relevance of the computerized recommendations was associated with an increase in implementation of the recommendations by the users.<sup>212</sup> In the study of Segal et al, a new CDSS was compared to a legacy CDSS and this showed higher clinical relevance (85% vs 16%), lower alert burden (0.4% vs 37%) and more changes in practice (43% vs 5%) for the new CDSS.<sup>207</sup> These outcomes are in line with a systematic review on factors influencing the uptake of medication-related CDSSs in hospitals. That review reported that the most frequent barriers were too many irrelevant alerts and information being not trustworthy.<sup>238</sup>

However, a lower clinical relevance of alerts may be accepted when an alert prevents severe adverse events.<sup>41</sup>

In 33% of the studies included in our review, the output of the CDSS was compared with a chart/medication review in the whole study population and therefore a CDSS coverage could be calculated (group C). This percentage is comparable to the systematic review of Foster et al, in which nine out of 24 (38%) studies compared the accuracy of ADE detection rules to a chart review in the whole study population.<sup>200</sup> The reported sensitivity (28-85%) and specificity (42-75%) in our review are within or near the broad range found by Foster et al of 40-94% (Sn) and 1.4-89% (Sp). Most studies in our scoping review (17/30) described step 3 (prospective validation) of the validation strategy of Scheepers-Hoeks et al.<sup>41,199</sup> We reviewed clinical validation studies and therefore step 1 (technical CDSS validation) of the strategy was not within the scope of our review. None of the included studies reported all steps (2, 3 and 4) of the validation strategy and none referred to a published CDSS validation strategy. Furthermore, only one study described the expected or desired results (threshold), beforehand. Future studies would benefit from following a more structured and comprehensive validation strategy and including a pre-defined threshold for either PPV, NPV, Sn, or Sp.

A strength of our study is that we used a systematic approach with a broad search conducted in two databases and that the screening and data extraction was conducted by two researchers. Another strength is the use of a risk of bias assessment tool. We used an adapted version of the QUADAS-2 tool and two of the included studies had one domain flagged with a high risk of bias. Validation methods and outcomes not described in the studies were not included in the narrative analysis, which may have resulted in an underrepresentation of the used methods and outcomes.

In conclusion, 30 articles validating a CDSS to support (part of) a medication review were found. These studies varied markedly in methods and outcome measures. Most included studies described the clinical relevance of the alerts and only 33% described a comparison of CDSS's output with all DRPs in the whole population. The studies did not report the use of a CDSS validation strategy. For the validation of their CDSSs, future studies would benefit from following a systematic and comprehensive approach as described in Scheepers-Hoeks et al, in order to limit irrelevant information, limit alert fatigue and, therefore, improve the uptake of CDSSs supporting (part of) a medication review.

