

Jak číst kriticky odborné články

^{1,2}Pavel Mohr, ^{1,2}Cyril Höschl, ³Jan Volavka

¹Psychiatrické centrum Praha

²3. lékařská fakulta, Univerzita Karlova Praha

³New York University School of Medicine

SOUHRN

Východisko. Snaha udržet si přehled v oboru je komplikována enormním nárůstem dostupných informací, ve kterých je obtížné se orientovat. Psychiatři i ostatní kliničtí lékaři mají tendenci číst odborné články povrchně a často spoléhají pouze na abstrakta. Autoři se pokusili zlepšit tuto situaci pomocí výuky systematického kritického hodnocení konkrétních publikací.

Metody. Sestavili jsme 23položkový hodnotící nástroj k posouzení různých aspektů psychofarmakologických článků. Hodnotící formulář byl použit k hodnocení konkrétních článků v průběhu workshopů s mladými psychiatry. Výsledky byly zaznamenány pomocí elektronického hlasovacího systému a následně diskutovány se zkušenými lektory, kteří také poskytli klíčové hodnocení každé položky.

Výsledky. Tři workshopů se zúčastnilo celkem 58 účastníků, diskutováno bylo šest publikací. Hodnocení článků přineslo heterogenní výsledky, které odrážejí rozdíly v teoretické přípravě účastníků i rozdílnou kvalitu publikací. Jako ilustrativní příklad prezentujeme podrobnou analýzu jednoho z článků.

Závěry. Naše zkušenosti ukázaly, že diskuzní formát a hlasovací systém s bezprostřední zpětnou vazbou stimuluje aktivní zapojení účastníků. Použití strukturovaného hodnotícího nástroje, který lze snadno použít i pro jiné odborné publikace, může zlepšit pochopení odborných textů a jejich kritické zhodnocení.

Klíčová slova: lékařské informace, publikace, psychofarmakologie, výuka.

SUMMARY

Mohr P, Höschl C, Volavka J. How to read critically a scientific paper

Background. Effort to follow the development in various fields of medicine is being hindered by enormous growth of information output. The vast amount of data makes it difficult to figure out what is really important. Psychiatrists and other physicians tend to read scientific papers superficially, relying excessively on abstracts. We addressed this problem by teaching critical appraisal of individual articles.

Methods. We developed a 23-item appraisal instrument to assess various aspects of papers on psychopharmacology. The appraisal form was used to evaluate individual articles during a series of workshops with psychiatric trainees. The results were collected with an electronic voting system and subsequently discussed with experienced tutors who also provided key ratings.

Results. The total of 58 trainees participated at three workshops, six publications were analyzed. Evaluation of the papers yielded heterogeneous results reflecting variations of the participants' theoretical background as well as varied quality of the publications. We present detailed analysis of one paper as an illustrative example.

Conclusions. Our experience showed that the discussion format and the voting system with immediate feedback stimulate active involvement of the participants. Administration of the structured appraisal instrument that can be easily adapted for other medical publications may enhance understanding and critical appraisal of scientific texts.

Key words: medical information, publications, psychopharmacology, teaching.

Mo.

Čas. Lék. čes. 2012; 151: xx-xx

ÚVOD

Ti lékaři v klinické praxi, kteří se snaží udržet krok s aktuálním děním v oboru, mají k dispozici nepřeberně množství informací – v tištěné podobě i na webu. O tom, že počet publikací trvale roste, svědčí například analýza publikačního trendu v lékařské informatice v průběhu 20 let, mezi lety 1987–2006, podle které počet článků stoupá ročně o 12 % (1). Vzrůstá i objem informací z jednotlivých subspecializací psychiatrie, například počet odborných článků publikovaných ústavu dětské

a adolescentní psychiatrie/psychologie indexovaných v databázi PubMed se zvedl ze 491 v roce 2005 na 639 v roce 2010, což představuje nárůst o 30 % (2).

Pro časově vytížené lékaře může být snaha udržet si přehled komplikována nejen záplavou informací, ale také někdy obtížnou orientací v tom, co je skutečně relevantní, a to i v renomovaných periodikách. Problémem je také mnohdy odlišit objektivní hodnocení od skryté inzerce, a to i v případě kdy je deklarovaný konflikt zájmů autora (3). Mnoho lékařů automaticky věří tomu, co čte, zejména pak pokud jsou pod článkem podepsané respektované autority v oboru. Není rovněž bez zajímavosti, že míra důvěry v tištěné slovo nekoreluje ani se čtenářovým vzděláním. Nedostatek času, množství informací a spoléhání se na renomované autory přispívá k tendenci číst odborné texty povrchně. V tomhle ohledu jsou důležité zejména abstrakta článků, neboť to je často jediná část publikace, která je detailněji studována.

Obsah abstraktu, stejně tak jako to, co je v něm napsáno i co je zamlčeno, může podléhat nejrůznějším externím vli-

ADRESA PRO KORESPONDENCI:

doc. MUDr. Pavel Mohr, Ph.D.
Psychiatrické centrum Praha
Ústavní 91, 181 03 Praha 8
fax: +420 266 003 366
e-mail: mohr@pcp.lf3.cuni.cz

vům. Vliv sponzora a jeho efekt na text článku, zejména pak abstraktu, byl dobře popsán (4). Jako konkrétní příklad může posloužit publikace srovnávající aripiprazol s haloperidolem (5). V souhrnu se zdůrazňuje lepší účinnost aripiprazolu, ale již nikde se nezmiňuje, že v žádném předem definovaném kritériu účinnosti rozdíl mezi léky nedosáhl statistické význam-

nosti. Autoři navíc provedli velké množství testů, ale neprovedli korekci této multiplicity, což je běžný problém, se kterým se lze v literatuře setkat.

Abstrakta mohou být zavádějící i v případě, kdy není žádný vnější vliv zjevný. Tak například se lze dočíst, že „první dva týdny léčby se doporučují střední dávky haloperidolu, v při-

Tab. 1. Hodnotící formulář s příkladem výsledků jednoho hodnocení (článek autorů Jones et al., 2006). Data označují kumulativní hlasování pro každou položku a shodu s klíčovým hodnocením.

Položka	Klíčové hodnocení ¹			Hodnocení účastníků (N = 27 ²)			Shoda s klíčovým hodnocením (% správných odpovědí)
	souhlasím	nesouhlasím	nevím	souhlasím	nesouhlasím	nevím	
1. Byla jasně definována hypotéza?	1			24	0	2	92
2. Byla kritéria pro zařazení pacientů jasná a adekvátní ke stanovené hypotéze?	1			13	10	3	50
3. Byl plánovaný počet pacientů stanoven analýzou síly testů (power analysis)?	1			22	3	1	85
4. Bylo na konci studie jasně popsáno, co se stalo s každým konkrétním pacientem, který vstoupil do studie?	1			24	0	2	92
5. Byla provedena intent-to treat analýza?	1			24	1	1	92
6. Byla provedena per protocol analýza?	1			21	4	1	81
7. Poradily si analýzy adekvátně s pacienty, kteří předčasně vypadli ze studie?	1			4	7	15	15
8. Kromě studijní léčby, byly skupiny pacientů léčeny stejně?			1	2	4	21	78
9. Byly porovnávané léky podávány ve srovnatelných dávkách?		1		5	19	3	70
10. Jak byla měřena compliance pacientů se studijní léčbou?	1			17	3	7	63
11. Byla randomizace adekvátní?	1			25	0	2	93
12. Bylo zaslepení studie dostatečné (ze strany pacientů, hodnotitelů a kliniků)?		1		10	13	3	48
13. Byly hodnotící metody vhodné pro testování stanovené hypotézy?	1			23	0	4	85
14. Byla dostatečná shoda mezi hodnotiteli v hodnocení použitých škál?	1			20	0	7	74
15. Byla plánovaná délka studie dostatečná?	1			12	4	11	44
16. Testovaly statistické analýzy hypotézu vhodným způsobem?	1			10	0	17	37
17. Testovaly statistické analýzy statistickou významnost?	1			23	2	1	88
18. Zhodnotily statistické analýzy velikost efektu (effect size)?	1			19	3	5	70
19. Pokud byly provedeny mnohočetné testy, byla provedena korekce?			N/A	5	3	19	70 ³
20. Byly analýzy přídatné medikace dostatečné?		1		0	26	1	96
21. Byly náležitě ošetřeny etické aspekty studie, včetně ochrany pacientů?	1			16	5	6	59
22. Jsou závěry autorů v souladu s jejich daty?	1			13	8	6	48
23. Jiné problémy ve studii a komentáře							

¹ „klíčové hodnocení“ označuje odpovědi lektorů, ² pro položky 1–7 a 17 je N = 26, ³ „nevím“ je započítáno jako „nelze aplikovat“

padě nedostatečné odpovědi zvýšení dávky nepřineslo žádný další prospěch“ (6). Později se však dozvídáme, že průměrné navýšení dávky bylo pouze o 0,72 mg týdně, což vsutku není klinicky významná změna. Je zřejmé, že zevšeobecnující tvrzení z abstrakt je třeba brát s nezbytnou rezervou. Shodou okolností tento citovaný článek je rovněž příkladem situace, kdy není k dispozici dostatek informací o studijní populaci, což znesnadňuje jakékoliv pokusy o replikaci nebo generalizaci výsledků.

V psychofarmakologických publikacích existuje celá řada dalších metodologických problémů. Některé z nich, včetně konkrétních příkladů, jsou uvedeny ve výše citovaném přehledu Herese et al. (4). Vedle zmiňovaného sponzorství jsou výsledky a jejich interpretace snadno ovlivněny například výběrem pacientů do studie a vstupními kritérii, zvoleným dávkováním a způsobem titrace medikace, použitými statistickými metodami nebo způsobem prezentace výsledků, volbou formulace v textu a publikačním bias (multiplicita publikací, prezentace formou posterů apod.).

Povrchní čtenář si těchto problémů není vždy vědom, obzvlášť pokud čte pouze souhrny článků. Podle naší zkušenosti mnoho lékařů v praxi nepovažuje tyto problémy za důležité, považují je spíše za „zanedbatelné technické detaily“. V tomto článku popisujeme naši snahu zlepšit tuto situaci pomocí výuky systematického kritického hodnocení konkrétních publikací. Jedná se o problém společný všem odborným publikacím, pro potřeby konkrétního experimentu jsme zvolili práce z psychofarmakologie. Zkrácená verze článku byla publikována jinde (7).

METODY

V letech 2009 a 2010 autoři uskutečnili na třech národních psychiatrických setkáních v České republice a na Slovensku sérii workshopů, na kterých učili kritickou analýzu psychofarmakologických publikací. Účastníci byli psychiatři ve specializační přípravě a další mladí kliničtí psychiatři z obou zemí. U každé výukové lekce byli přítomni lektori se zkušeností s klinickými studii (včetně autorů), kteří řídili diskusi a poskytovali účastníkům zpětnou vazbu. Na každém workshopu se analyzovaly dva různé články, které měli účastníci k dispozici předem, aby si je mohli individuálně prostudovat. Použitým nástrojem byl hodnotící formulář s 23 položkami, který navrhli autoři. Formulář částečně vychází z obdobného nástroje, který byl vyvinut v Center for Evidence-Based Medicine na University of Toronto (<http://ktclearinghouse.ca/cebm/teaching/worksheets/therapy>). Dvacet dva položek formuláře jsou otázky, které zkoumají různé aspekty článku, se třemi možnými odpověďmi: „souhlasím“, „nesouhlasím“, „nevím“. Dodatečně byla doplněna čtvrtá možnost („nelze aplikovat“). Poslední doplňková položka je nestrukturovaná otázka, která se dotazuje na další problémy studie a dává možnost komentáře. Plný text hodnotícího formuláře s příklady výsledků je uveden v tabulce 1.

V průběhu workshopu každý účastník hodnotil články individuálně položku za položkou pomocí hodnotícího formuláře. Výsledky byly sbírány a zaznamenávány anonymně pomocí elektronického hlasovacího systému. Text každé položky byl promítnut na projekční plátno a účastníci pak samostatně zvolili jednu ze čtyř odpovědí („souhlasím“, „nesouhlasím“, „nevím“, „nelze aplikovat“) stisknutím tlačítka na hlasovacím zařízení, které dostal každý k dispozici. Poté co všichni účastníci odhlasovali, se na projekčním plátně okamžitě zobrazily výsledky ukazující četnost jednotlivých odpovědí na hodnocenou otázku a současně se výsledky zaznamenaly pro pozdější analýzu. Poté následovala diskuze nad konkrétní položkou a její hodnocení. Lektori se podělili o svoje názory a poskytly zpětnou vazbu, včetně správných (klíčových) odpovědí.

Pro ilustraci jsme jako příklad vybrali hodnocení jednoho článku ze setkání, na kterém byl nejvyšší počet frekventantů. Hodnotili jsme následující proměnné:

1. položky s nejvyšší shodou s klíčovým hodnocením, měřeno jako procento „správných“ odpovědí,

2. položky s nejvyšší neshodou mezi účastníky, měřeno jako frekvence nejednotných (diskordantních) odpovědí,

3. položky s nejvyšším počtem nerozhodných odpovědí („nevím“), kde však přitom klíčové hodnocení bylo „souhlasím“ nebo „nesouhlasím“.

VÝSLEDKY

V průběhu tří workshopů bylo hodnoceno celkem šest článků (5–6, 8–11). Celkový počet školících se účastníků byl 58. Metodologická kvalita stejně tak jako srozumitelnost se mezi články velmi lišila. K rozboru jsme zvolili hodnocení článku Jonese a spolupracovníků (10), jehož výsledky jsme dále analyzovali (tab. 1).

Výsledky ukázaly, že:

1. Nejvyšší shody s klíčovým hodnocením (více než 90 % „správných“ odpovědí) bylo dosaženo v položkách hodnotících, zda byla analýza přídatné medikace dostatečná (96 %), adekvátnost randomizace (93 %), zda byla hypotéza jasně definovaná (92 %), zda bylo jasně popsáno, co se stalo s každým konkrétním pacientem, který vstoupil do studie (92 %), zda byla provedena intent-to-treat analýza (92 %). Zdaleka nejnižší shoda (15 %) byla nalezena v položce hodnotící adekvátnost analýzy pacientů, kteří ukončili studii předčasně. Nízká shoda s klíčovým hodnocením (méně než 50 % „správných“ odpovědí) byla také zaznamenána v položkách hodnotících, zda statistické analýzy testovaly hypotézu vhodným způsobem (37 %), zda byla plánovaná délka studie dostatečná (44 %), zda bylo zaslepení dostatečné (48 %) a zda jsou závěry autorů v souladu s jejich daty (48 %).

2. Největší neshoda mezi účastníky, kdy byla četnost jednotlivé odpovědi (bez ohledu zda „správné“, nebo „nesprávné“) nižší než 50 %, byla zaznamenána v položkách hodnotících kritéria pro zařazení pacientů (souhlasím 50 %, nesouhlasím 38 %, nevím 12%), zaslepení (38–50–12 %), délku studie (44–15–41 %) a zda jsou závěry autorů v souladu s jejich daty (48–30–22 %).

3. Nejvyšší počet odpovědí „nevím“ udávali účastníci v položkách hodnotících adekvátnost statistických metod testujících hypotézu (63 %), adekvátnost zacházení s předčasně ukončenými pacienty (58 %) a plánovanou délku studie (41 %).

DISKUZE

Sestavili jsme nástroj k hodnocení psychofarmakologických článků a použili jsme jej jako výukový nástroj v sérii lekcí pro mladé psychiatry. Hodnotící formulář zdůrazňuje základní metodologické charakteristiky, které jsou důležité pro interpretaci výsledků psychofarmakologických studií. Hodnotící formulář byl použit k hodnocení konkrétních publikací.

Primárním účelem celého výukového experimentu byla pomoci lékařům ve specializační přípravě k tomu, aby se naučili přemýšlet kriticky o tom, co čtou. Aplikací systematického přístupu pomocí seznamu položek hodnotícího formuláře mohli sami objevit různé metodologické problémy v publikacích, které byly v průběhu workshopů podrobeny zkoumání. Jednou z hlavních úloh lektorů bylo ukázat, jaký dopad mohou mít tyto problémy na interpretaci výsledků. Diskuzní formát a hlasování stimulovaly aktivní zapojení účastníků. Zaznamenali jsme, že aktivní účast usnadňuje strukturovaným hodnotícím nástrojem, s následnou bezprostřední zpětnou vazbou a myšlenky stimulující diskuze, mohou posílit proces učení. V neposlední řadě naše výsledky zdůrazňují význam výuky základů statistiky a metodologie.

Hodnotící nástroj byl specificky zkonstruován k ocenění

článků týkajících se farmakologických studií. Jednotlivé položky jsou však snadno aplikovatelné i na jiné medicínské práce, ať už klinické, či preklinické studie. S minimální modifikací lze tedy hodnotící formulář použít i ke kritickému čtení další odborné literatury a lepší orientaci v oboru.

Naše předběžná studie, jež předkládá zkušenosti s touto novou výukovou metodou, má svá omezení. Vzhledem k malému počtu hodnocených článků a počtu účastníků nebyla možná žádná hlubší statistická analýza. Výsledky, které zde prezentujeme, slouží jen jako ilustrační příklad a neumožňují žádné široké zobecnění. Je třeba také upozornit na fakt, že jsme účastníky netestovali před a po absolvování workshopů, abychom mohli lépe odhadnout edukační dopad experimentu. Všechny tyto problémy lze ošetřit v dalším výzkumu.

LITERATURA

1. **Deshazo JP, et al.** Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH. *BMC Med Inform Decis Mak* 2009; 9: 7.
2. **Albayrak O, et al.** Bibliometric assessment of publication output of child and adolescent psychiatric/psychological affiliations between 2005 and 2010 based on the databases PubMed and Scopus. *Eur Child Adolesc Psychiatry* 2012; 21: 327–337.
3. **Pus PE.** Vliv farmaceutického průmyslu na vzdělávání lékařů. *Psychiatrie* 2004; 8: STRANY???
4. **Heres S, et al.** Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry* 2006; 163: 185–194.
5. **Kasper S, et al.** Efficacy and safety of aripiprazole vs. haloperidol for long-term maintenance treatment following acute relapse of schizophrenia. *Int J Neuropsychopharmacol* 2003; 6: 325–337.
6. **Giegling I, et al.** Interaction of haloperidol plasma level and antipsychotic effect in early phases of acute psychosis treatment. *J Psychiatr Res* 2010; 44: 487–492.
7. **Mohr P, et al.** Teaching critical appraisal of articles on psychopharmacology. *Acad Psychiatry* 2012; 36: 114–117.
8. **Chengappa KN, et al.** Clozapine: its impact on aggressive behavior among patients in a state psychiatric hospital. *Schizophr Res* 2002; 53: 1–6.
9. **Dose M, et al.** Treating schizophrenia patients with antipsychotics in a naturalistic setting: a retrospective chart review. *Psychiatrie* 2008; 12: 197–200.
10. **Jones PB, et al.** Randomized Controlled Trial of the Effect on Quality of Life of Second- vs First-Generation Antipsychotic Drugs in Schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry* 2006; 63: 1079–1087.
11. **Kahn RS, et al.** Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. *Lancet* 2008; 371(9618): 1085–1097.

Autoři děkují společnosti Eli Lilly za podporu workshopů.