



**Az emlőrák prognosztikai tényezői
reprodukálhatóságának ONEST (Observers Needed
to Evaluate Subjective Tests) vizsgálata**

Ph. D. Tézis

CSERNI BÁLINT, M.Sc.

**Témavezető:
Cserni Gábor, M.D., D.Sc.**

**Multidiszciplináris Orvostudományok Doktori Iskola
Szegedi Tudományegyetem
Szeged, Magyarország**

2023

A PH.D. TÉZIS ALAPJAIUL SZOLGÁLÓ PUBLIKÁCIÓK LISTÁJA

I. **Cserni B**, Bori R, Csörgő E, Oláh-Németh O, Pancsa T, Sejben A, Sejben I, Vörös A, Zombori T, Nyári T, Cserni G. The additional value of ONEST (Observers Needed to Evaluate Subjective Tests) in assessing reproducibility of oestrogen receptor, progesterone receptor and Ki67 classification in breast cancer. *Virchows Arch* 2021; 479(6):1101-1109. doi: 10.1007/s00428-021-03172-9

IF(2021): 4.535 (Scimago journal ranking: Q1)

II. **Cserni B**, Bori R, Csörgő E, Oláh-Németh O, Pancsa T, Sejben A, Sejben I, Vörös A, Zombori T, Nyári T, Cserni G. ONEST (Observers Needed to Evaluate Subjective Tests) suggests four or more observers for a reliable assessment of the consistency of histological grading of invasive breast carcinoma - A reproducibility study with a retrospective view on previous studies. *Pathol Res Pract* 2022;229:153718. doi: 10.1016/j.prp.2021.153718.

IF(2021): 3.309 (Scimago journal ranking: Q2)

ONEST PROGRAM, MELY A PH.D. TÉZISHEZ LETT FEJLESZTVE

Cserni B. ONEST Calculator. <https://github.com/csernib/onest>

1. BEVEZETÉS

A bőr rosszindulatú daganatai kivételével az emlőrák a leggyakoribb rosszindulatú tumor a nők körében Európában és a világ a többi részén.

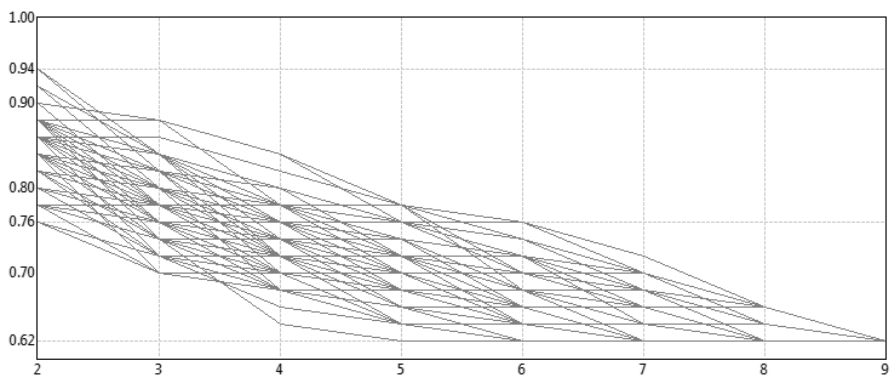
Az emlőrák több prognosztikai tényezőjének meghatározása hagyományos szövettani festéssel (hematoxinin és eozin, HE) vagy immunhisztokémiával (IHC) történik, azonban a paraméterek értelmezése szubjektív elemeket tartalmaz, és így függ a vizsgálok sokszínűségétől. Ez a doktori tézis az alább részletezett prognosztikai tényezők reprodukálhatóságának bizonyos aspektusaival foglalkozik.

Az emlőrák osztályozásánál az egyik legfontosabb elem a daganatok ösztrogén receptor pozitív (ER+), illetve negatív (ER-) csoportokba való besorolása, ugyanis jellemzően csak az ER+ esetekben hatékonyak az endokrin kezelések. Jelenleg az ER státusz általánosságban IHC-vel kerül meghatározásra, és az ER pozitivitás eldöntése valamilyen szinten önkényesen történik. A progesteron receptorok (PR) szintén hatással vannak az endokrin kezelés hatásosságára, és meghatározásuk az ER-hez hasonló módon történik. A Ki67 egy fehérje, amely a G0 fázis kivételével változó mértékben jelen van a sejtciklus egésze alatt, és így a sejtosztódás prognosztikai szempontból fontos jelzője. Az ER, PR és Ki67 mikroszkóppal történő értékelése a megfelelő antitestekkel festődő sejtmagok számszerűsítéséből áll, melynek egy gyakori módszere a festődött tumorsejtek számának vizuális úton történő becslése.

A differenciáció foka (grade) a daganat biológiai viselkedésére utaló prognosztikai paraméter. Az értéke három pontszám összegéből adódik, melyeket a tubulusképzés, magpleomorfizmus és mitotikus aktivitás alapján adnak. A grade prognosztikai jelentősége ellenére nem tökéletesen reprodukálható paraméter, aminek tényével több publikáció is foglalkozott már.

A fenti prognosztikai tesztek karakterizálásához, mint új megközelítést, a nemrég kifejlesztett ONEST (Observers Needed to Evaluate Subjective Tests, azaz Szubjektív Tesztek Kiértékeléséhez Szükséges Vizsgálók) módszert

alkalmaztuk. Az ONEST a patológusok (vizsgálók) száma függvényében vett OPA (Overall Percent/Proportion Agreement, vagyis a teljes egyezés százaléka/aránya) értékek ábrázolásán alapul. A vizsgálók összes lehetséges permutációja közül véletlenszerűen kiválasztunk 100-at, melyek mindegyikéhez egy olyan függvényt rendelünk, amely a permutáció minden i . eleméhez a permutációban i -nél nem nagyobb helyen álló vizsgálók közötti OPA értéket rendeli. Ezeket a függvényeket ábrázolva OPA görbéket (OPA curves, OPAC) kapunk, és a 100 görbe együttese adja az ONEST ábrát. Jól reprodukálható tesztek magas $OPA(n)$ -t (azaz az összes vizsgálóra nézett OPA értéket) adnak, kevés vizsgálóval elérik a platót, ahonnan az OPAC-ok már nem csökkennek (ONEST érték) és kis különbséggel rendelkeznek a legjobb és a legrosszabb OPA között (sávszélesség).



A fenti ábra példaként a Ki67-tel festett biopsziák (CNB) $<1\%$, $1-10\%$ és $>10\%$ kategorizálású ONEST görbéit mutatja. Az X tengelyen a vizsgálók száma, az Y-on az OPA értékek láthatók. Az ábrát 100 (gyakran egymást átfedő) OPAC alkotja. Az alsóértékek 5 vizsgálónál érik el a platót (ONEST érték), vagyis a 0,62 OPA-t, amely egyben az $OPA(9)$ is. A sávszélesség $0,94 - 0,76 = 0,18$, azaz a legnagyobb két vizsgáló közötti különbség 18%.

Az ONEST eredeti leírása alapján a szerző egy számítógépes eszközt készített, mely segít az ONEST számítások elvégzésében.

2. CÉLKITŰZÉSEK

Egy univerzális számítógépes program fejlesztése ONEST számításokhoz, illetve annak használata a szükséges vizsgálószám becslésére az emlőrák olyan prognosztikai és prediktív tényezői reprodukálhatóságának megbízható számításához, mint az ER, PR vagy Ki67 IHC alapján történő kiértékelése (I. tanulmány), illetve a grade és komponenseinek HE festett metszetek alapján történő pontozása (II. tanulmány).

Korábbi reprodukálhatósági vizsgálatok elemzése az ONEST eredményeinek tükrében.

3. ANYAG ÉS MÓDSZEREK

A Bács-Kiskun Megyei Oktatókórház archívumából kiválasztottunk 100 emlőrák esetet, melyeken korábban rutin HE, valamint ER, PR és Ki67 IHC festést végeztek. Az esetek 50 biopsziás (CNB) és 50 excíziós (EXC) mintát tartalmaztak, egymástól független daganatokból.

A résztvevők feladata az volt (I. tanulmány), hogy jegyezzék le a festődő daganatos sejtek arányát mindhárom IHC reakciónál, az ER és PR esetében hozzávéve azok festődési intenzitását és Allred pontszámát. Az ER és PR százalékokat ezután három osztályba soroltuk: negatívok (<1%), gyengén pozitívok (1-10%) és pozitívok (>10%). A Ki67 értékek az emlőrákra vonatkozó magyar patológiai ajánlást követve, szemrevételezés alapján kerültek becslésre, mindig a legközelebbi 5%-hoz kerekítve az eredményt. Ki67 esetében öt kategorizálást vizsgáltunk: (1) ugyanazokkal a határértékekkel, mint az ER és PR

esetében – bár ezt a kategorizálást gyakorlatban nem használják, lehetővé teszi a közvetlen összehasonlítást az ER/PR értékekkel; (2-5) a 2009-es, 2011-es, 2013-as és 2015-ös St. Galleni Konszenzus Konferenciákon ajánlott határértékek alapján. Az értékelés megbízhatóságát osztályon belüli korrelációs együtthatóval (intra-class correlation coefficient, ICC) analizáltuk.

Ezzel párhuzamosan (II. tanulmány), az emlőrákok WHO szerinti legfrissebb osztályozását követve, mindegyik vizsgáló elvégezte a 100 eset gradálását is, pontozva a grade összetevőit (tubulusképzés, magpleomorfizmus és mitotikus aktivitás), illetve megadva magát a grade-et is. A grade reprodukálhatóságának vizsgálatára leíró statisztikát, illetve ICC-t és Fleiss-féle kappa értékeket használtunk.

Az ONEST-et, Reisenbichler et al eredeti leírása szerint (*Mod Pathol 2020; 33: 1746-1752*), a 362 880 (=9!) lehetséges patológus-permutációból kiválasztott 100 permutációra alkalmaztuk. A Kruskal–Wallis statisztikai próbát használtuk a minimum görbék összehasonlítására. Itt a $p < 0,05$ értékeket tekintettük statisztikailag szignifikánsnak. A statisztikai próbákhoz a Real Statistics Resource Pack Excel add-int használtuk.

Az eredményeink alapján a grade korábbi reprodukálhatósági tanulmányait is megvizsgáltuk, és azokat statisztikai módszerük, illetve a használt vizsgálószámuk alapján analizáltuk.

4. EREDMÉNYEK

A 100 eset kiértékelését 9 patológus, köztük 2 emlőpatológiában jártas rezidens végezte. A vizsgálok mindegyike több, mint 1 év, a leggyakorlottabb több, mint 25 év emlőpatológiai tapasztalattal rendelkeztek.

A vizsgált paraméterekre számított ICC eredmények alapján a legtöbb ER és PR státusszal kapcsolatos osztályozás (százalékos és Allred pontszám) kiváló

vagy jó-kiváló megbízhatósággal rendelkeznek. Ezzel szemben mindegyik Ki67-tel kapcsolatos osztályozás mérsékelt vagy mérsékelt-jó megbízhatóságú. Az ICC értékek közötti különbség szembetűnő az 1% és 10% küszöbértékekkel vett háromkategóriás osztályozás esetén az ER/PR (ICC: 0,909-0,996) és Ki67 (ICC: 0,625 EXC; 0,673 CNB) között. A Ki67 kategorizálások közötti ICC értékek különbsége kevésbé feltűnő (mindegyik $\leq 0,760$). Alapvetően nincs számottevő vagy konzisztens különbség a CNB és EXC minták ICC értékei között.

Az $<1\%$, $1-10\%$ és $>10\%$ küszöbértékeket használva jelentős különbségek voltak az ONEST ábrákról leolvasott minimum OPA értékek között mind a CNB, mind az EXC esetekben, bármelyik két ER, PR vagy Ki67 ábrát is vetettük össze.

A Ki67 festődési indexek egymást követő St. Gallen konszenzus konferenciák ajánlása szerinti alacsony vagy magas (vagy közepes, ha ezt is definiálták) proliferációba történő kategorizálásával kapcsolatban megállapítottuk, hogy a legnagyobb OPA a 2013-as ajánlással mutatkozott, vagyis a 20%-os küszöbérték használatával. Ez jelentősen jobb eredményt adott, mint bármelyik másik St. Gallen ajánlás, ugyanakkor az ICC értékek továbbra is csak mérsékelt-jó (CNB), illetve jó (EXC) reprodukálhatóságról tanúskodtak.

Mivel a 9! (362 880) még egy kezelhető szám, a 100 véletlenszerű permutációból vett vizsgálószám szerinti minimum OPA értékeket összehasonlítottuk az összes permutációból vett vizsgálószám szerinti minimum OPA értékekkel, azaz a legalacsonyabb OPAC-cal. Nem találtunk szignifikáns eltérést; a legtöbb esetben a Kruskal–Wallis próba $p=1$ -et adott, de a legkisebb p érték is 0,64 volt.

Az ONEST érték (vagyis a reprodukálhatóság megbízható becsléséhez szükséges vizsgálószám) 2 volt az ER és 3 PR esetén, mind a CNB, mind az EXC mintákon. Ki67-nél az ONEST érték 4-6 között volt.

A hisztológiai grade-re alkalmazott kappa és ICC statisztika eredménye mérsékelt, illetve mérsékelt-jó reprodukálhatóságot tükröz. A grade összetevőinek pontozása ennél gyengébben reprodukálható, és ezen belül is a

magpleomorfizmusnál a legrosszabb, míg a tubulusképzésnél a legjobb. Érdekes módon a tubulusképzés és a magpleomorfizmus is konzisztensebben lett pontozva az EXC mintákon. Általánosságban elmondható az is, hogy a középkategóriák (pontszámok) kevésbé voltak reprodukálhatók, mint a szélsőértékek.

Az ONEST analízis eredménye szerint minimum 4 patológus szükséges a grade reprodukálhatóságának megbízható értékelésére. Ez az a pont, ahol a minimum OPAC-ok meredeksége látszólag csökkenni kezd, míg végül CNB esetén 6, EXC-nél pedig 7 vizsgálószámnál érik el a platót.

A minimum OPA értékeket tekintve szignifikáns volt a különbség CNB és EXC esetek között magpleomorfizmus (Kruskal–Wallis, $p=0,006$) és grade ($p=0,042$) esetén. Az előbbinél az EXC, míg az utóbbinál a CNB ért el magasabb minimum OPA értékeket. A tubulusképzés és mitotikus aktivitás minimum görbéi között nem volt statisztikailag számottevő eltérés CNB és EXC minták között.

A grade reprodukálhatóságának kappa és összes vizsgálószámra értelmezett OPA értékeket használó korábbi tanulmányait is megvizsgáltuk. Az eredmények arra utalnak, hogy a reprodukálhatóság (virtuálisan) jobbnak tűnik, ha 4-nél (az ONEST értéknél) kevesebb vizsgáló vesz benne részt.

5. DISZKUSSZIÓ

Ismert, hogy az ER, PR és Ki67 IHC alapján történő értékelése sok tényezőtől függ. Ez a tanulmány csak az kiértékeléssel kapcsolatos nehézségekre koncentrált, habár két különböző típusú mintát is felhasznált: a nagyobb méretű excíziós mintákhoz képest a biopsziák jobban rögzülnek és kisebb területtel rendelkeznek, ami csökkentheti a különbségeket a vizsgálók között.

Az ER, PR és Ki67 összehasonlítása az 1% és 10%-os küszöbértékekkel arra utal, hogy a Ki67 a legkevésbé reprodukálható, ami valószínűleg a festődött sejtek arányának viszonylag széles skálájával magyarázható. Az ER/PR-hez képesti alacsonyabb ICC-vel egyetemben, az ONEST analízis is gyengébb

értékeket produkál Ki67-re: a legnagyobb eltérés két vizsgáló OPA értéke között tágabb (akár 34% is), az összes vizsgálóra vonatkozó OPA alacsonyabb (26%-os minimum), és a reprodukálhatóság megbízható becsléséhez is több vizsgáló szükséges (többnyire 5).

A Nemzetközi Ki67 Emlőrákban Munkacsoport (International Ki67 in Breast Cancer Working Group) által leírt javuló ICC eredmények alapján egyértelmű, hogy a Ki67 értékelésének összhangja standardizált leletezéssel javítható digitális képelemzés nélkül is, és a megbízható Ki67 értékelés felé az út a standardizáción keresztül vezet. Ez a tanulmány ugyanakkor nem a reprodukálhatóság növelését tűzte ki célul, hanem a reprodukálhatóságot mint alapadatot vette, és ezt egészítette ki az újonnan kifejlesztett ONEST módszerrel, melynek segítségével elemezte, hogy mit lehet hozzáadni prognosztikai és prediktív következményekkel járó biomarkerek reprodukálhatósági vizsgálataihoz. Alapfeltevésünkkel összhangban, az ONEST módszer képes kiegészíteni a vizsgálok közötti egyetértés hagyományos statisztikai eszközeit. Bizonyíthatja vagy egyszerűen vizualizálhatja, hogy egy biomarker a könnyű értékelésének és természetes eloszlásának köszönhetően megbízható (pl. esetünkben az ER: magasan elhelyezkedő OPAC görbék szűk sáv szélességgel). Hasonlóképpen képes kimutatni egy adott biomarker kiértékelésének nehézségét is (sok eltérés a vizsgálok véleménye között, azaz széles sáv szélesség az alsó és felső görbe között, valamint alacsony OPA érték az összes vizsgálóra vetítve). Ezeket a hasznos adalékokat az ONEST a módszer eredeti célján – vagyis a plató eléréséhez szükséges vizsgálószám meghatározásán – túl nyújtja. Ebben az összefüggésben bizonyos korábbi tanulmányok eredményei is megkérdőjelezhetők, ha azokat az ONEST által javasolt vizsgálószámnál kevesebb egyén bevonásával végezték.

A tanulmányunk során a hisztológiai grade meghatározásának több korábbi reprodukálhatósággal kapcsolatos megfigyelését is meg tudtuk ismételni. Az Egyesült Királyság emlőpatológiai külső minőségbiztosítási rendszerének nagy múltú tapasztalatával egyetértésben mi is arra jutottunk, hogy a grade

komponensei közül a tubulusképzés a legjobban, a magpleomorfizmus pedig a legrosszabbul reprodukálható. A középső kategóriák alapvetően kevésbé reprodukálhatók, mint a szélsőértékek (az alacsony és magas pontszámú kategóriák), és a közép kategóriák közül is a mitotikus aktivitás a legkevésbé reprodukálható. Az általunk végzett ONEST analízis minimum 4-7 vizsgálószámot javasol grade, illetve annak komponensei reprodukálhatóságának megbízható becsléséhez. Ennek fényében megvizsgáltunk az irodalomban fellelhető több korábbi OPA elemzést is, és megállapítottuk, hogy az OPA értékek látszólag jobbnak tűnnek azokban az tanulmányokban ahol a vizsgálatot 4-6 vizsgálónál kevesebben végezték. Kappa statisztikát használó korábbi tanulmányok szintén hasonló trendet követtek. Megjegyzendő, hogy több korábbi, a grade reprodukálhatóságával foglalkozó tanulmány a Cohen-féle kappát alkalmazta, amely mindössze két vizsgáló közötti egyezést mér. Ennek következtében páronkénti összehasonlítások történtek, és egy tartomány vagy átlag került publikálásra. Mivel ugyanakkor az így kapott eredmény két vizsgáló értékeléséből származtatott adatsoron alapul, a Cohen-féle kappa jobb eredményeket tükrözhet, mint ami az ONEST analízis alapján várható lenne.

A diskusszió végén érdemes kitérni néhány további gondolatra. A tumort infiltráló limfocitákkal (TIL) kapcsolatos, folyamatban lévő munkák és az ONEST módszer további analízisének során több olyan tényezőt is azonosítottunk, amelyek hatással lehetnek a tanulmányunk eredményeire.

Az ONEST ábrákról leolvasott következtetéseket befolyásolhatja a vizsgálók számának vagy tapasztalatának változtatása és a többségtől jelentősen eltérő véleményen lévő vizsgálók eltávolításával „javítható” az eredmény. Ez természetesen torzítja a valóság alapján megfogalmazható elvárásokat, hiszen a való életben sem rendelkezik minden vizsgáló egyforma szakmai hozzáértéssel, és ha a tényleges reprodukálhatóságot szeretnénk becsülni, akkor ezeket a divergens vizsgálókat sem szabad figyelmen kívül hagyni. Az I. és II. tanulmányunkban felismert tényezőkhöz kívül, mint amilyenek az osztályozáshoz használt kategóriák száma vagy az értékek szélsőértékekhez viszonyított

eloszlása, felismertük, hogy az eloszlásban lévő heterogenitás szintén hatással van az ONEST eredményeire, csakúgy, mint a reprodukálhatóság más mértékeire.

A tézis alapjait szolgáló publikációkban az ONEST értéket a minimum OPAC meredeksége alapján határoztuk meg, úgy tekintve, hogy a vízszintest erősen közelítő lejtő már kellő megközelítést tesz lehetővé az ONEST érték leolvasásához. A tézisben ezt módosítottuk úgy, hogy az approximáció mellett a plató tényleges elérését is figyelembe vettük. A publikusan elérhető szoftver szintén ennek fényében került kifejlesztésre.

6. KÖVETKEZTETÉSEK

Összefoglalva tehát, először alkalmaztuk az ONEST-et az ER, PR és Ki67 reprodukálhatóságának karakterizálására, melyek mindegyike a festett sejtmagok arányának becslésén alapul. Az egyes reprodukálhatósági mértékek közötti különbségek a festődött sejtmagok szélsőértékektől (0% és 100%) való eltérő eloszlásával magyarázhatók. Az ONEST módszer hasznos kiegészítő információt nyújtott, és segített vizualizálni az eredményeket. A minimális OPA értékek, a legnagyobb különbség két patológus OPA eredménye között (sávszélesség) és az összes vizsgáló alapján számolt OPA (OPA(n)) mind leolvasható az ONEST ábrákról.

A reprodukálhatóság megbízható becsléséhez szükséges vizsgálószám 2 volt az ER, 3 a PR és 4-6 között a Ki67 kategorizálásánál.

Az ONEST analízisünk alapján elmondható, hogy legalább 4, de inkább 6-7 vizsgáló szükséges a grade reprodukálhatóságának megbízható értékeléséhez. Ezzel összhangban a fenténél kevesebb vizsgálószámmal vagy páronkénti összehasonlítással végzett korábbi tanulmányok valamivel jobb egyezést mutatnak a grade meghatározásában, akár OPA, akár kappa értékeket nézünk. Az eredményeink összhangban állnak az olyan korábbi tanulmányokkal, amelyekben 3-nál több vizsgáló vett részt, és utalnak arra, hogy a grade mérsékelt vagy

mérsékelt-jó reprodukálhatósággal rendelkeznek. Mindezek mellett, a grade a többváltozós elemzésekből származtatott emlőrak prognosztikai eszközöknek továbbra is a része. A grade értékelésénél fellépő szórást el kell fogadni, de hatása csökkenthető tapasztalattal, a vizsgálónak küldött visszajelzéssel és elkötelezett kiértékeléssel.

Akárcsak más reprodukálhatósági mértéket, az ONEST-et is több olyan tényező befolyásolja, mint például a besorolási kategóriák száma, a vizsgált paraméter szélsőséges értékekhez közeli vagy távoli eloszlása, a paraméter homogenitása vagy heterogenitása, a vizsgálók tapasztalata és száma, az általános besorolástól következetesen eltérő vizsgálók jelenléte... stb. Ezek miatt az ONEST elemzések eredményét is becslés jellegűnek, a reprodukálhatóság elemzésére szolgáló kiegészítésnek kell tekinteni.

7. KÖSZÖNETNYILVÁNÍTÁS

Elsőként szeretnék köszönetet mondani édesapámnak és témavezetőmnek, Prof. Dr. Cserni Gábornak, akinek rendkívüli támogatása és útmutatása nélkülözhetetlen volt ennek a tézisnek a megírásához. Köszönöm, apa!

Köszönöm az összes társszerzőmnek, akik nélkül ez a tanulmány nem tudott volna létrejönni. Külön kiemelném Prof. Dr. Nyári Tibort, akinek statisztikában való jártassága kulcsfontosságú volt több számításunkhoz.

Külön köszönet illeti a Real Statistics Resource Pack Excel add-in szerzőjét, Charles Zaiontzt. Az eszköz, melyet készített, nagyon hasznosnak bizonyult a statisztikai számításainkhoz, és Zaiontz úr maga is személyesen segített statisztikai, illetve az eszközre vonatkozó kérdéseink megválaszolásában.

Az ONEST számításokhoz készített program fejlesztése nagyban épült a wxWidgets programkönyvtárra és a lest tesztelési keretrendszerre, melyek szerzőinek ezért külön hálás vagyok.

Hálás vagyok az összes tanáromnak. A Szegedi Tudományegyetemről szeretnék külön köszönetet mondani az Informatikai Intézetben dolgozó professzoraimnak, oktatóimnak és tanítóimnak, külön hangsúlyozva Dr. Nagy Antal, Dr. Ferenc Rudolf, Prof. Dr. Gyimóthy Tibor, Dr. Siket István és Dr. Vinkó Tamás munkásságát. Szeretném kifejezni a hálámat a tanárimnak a Kecskeméti Katona József Gimnáziumban, főként matematika tanáromnak, Dr. Szablics Bálintnének, és fizika/kémia tanáromnak, Sáróné Jéga-Szabó Irénnek. Az ő felkészítésük tette lehetővé, hogy az egyetemi tanulmányaim során fellépő első akadályokat könnyedén vegyem.

Köszönetem a családomnak és barátaimnak, akik mind segítettek a tézis megírásához szükséges út során.

A tézis alapjaként szolgáló cikkek megjelentetését a University of Szeged Open Access Fund támogatta (5440-es és 5580-as pályázat).